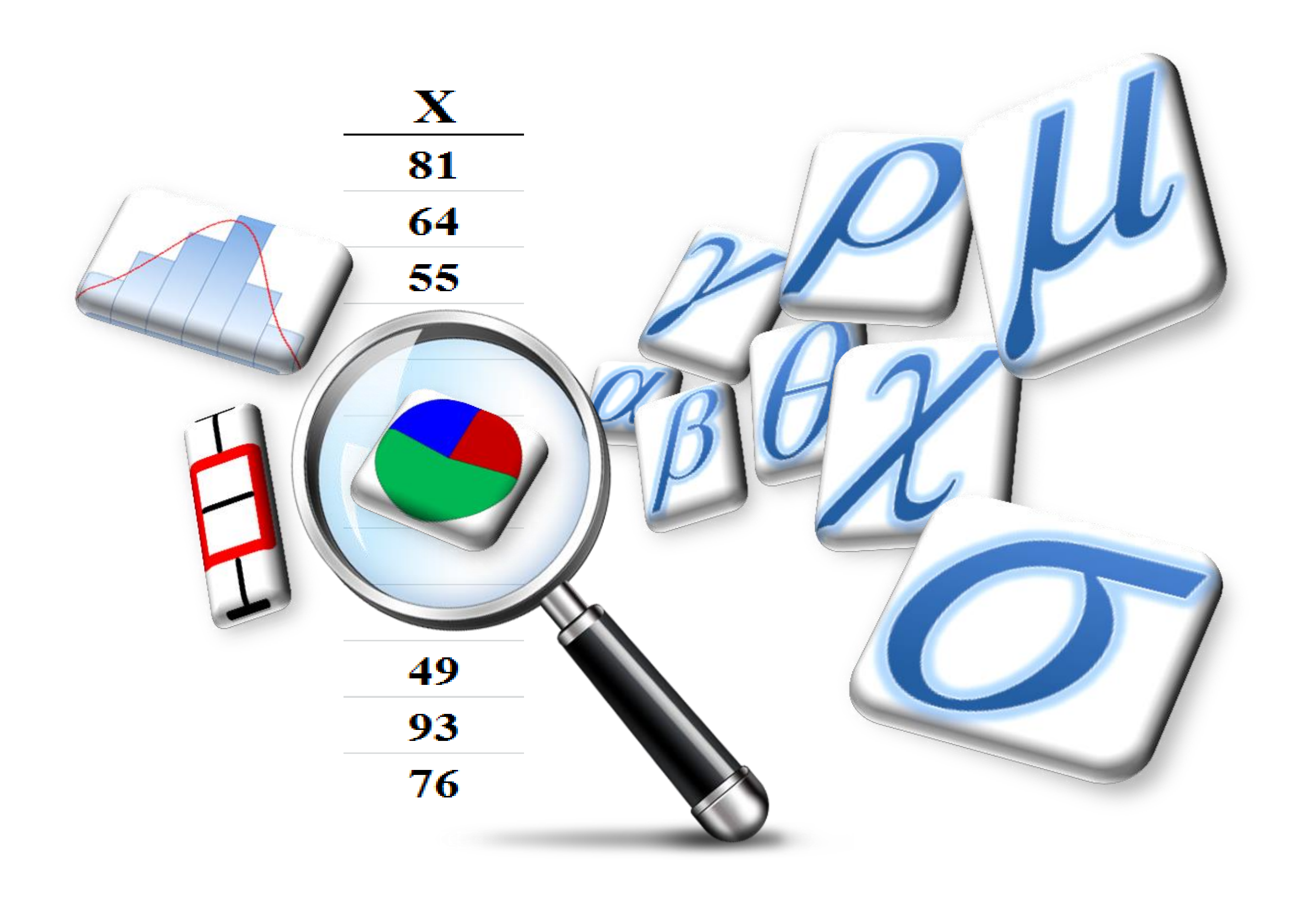
مقدمة في

الإحصاء الاستكشافي والاحتمالات

Introduction to Exploratory Statistics and Probability



د. رامي صلاح محمد جبريل

الطبعة الأولى – 2015م

مقدمة في

الإحصاء الاستكشافي والاحتمالات

(Introduction to Exploratory Statistics and Probability)

تأليف

د. رامي صلاح محمد جبريل

أستاذ مشارك في قسم الإحصاء – كلية العلوم

جامعة بنغازي – ليبيا

الطبعة الأولى – 2015م

الطبعة الأولى – 2015م

اسم الكتاب: مقدمة في الإحصاء الاستكشافي والاحتمالات

اسم المؤلف: د. رامي صلاح محمد جبريل

الوكالة الليبية للترقيم الدولي الموحد للكتاب

دار الكتب الوطنية

بنغازي – ليبيا

هاتف: 9097074 – 9096379 – 9090509

بريد مصور: 9097073

البريد الإلكتروني: nat_lib_libya@hotmail.com

ردمك ISBN 978-9959-1-1440-2

بسم الله الرحمن الرحيم

وبه نستعين

المقدمة

الحمد لله رب العالمين، والصلاة والسلام على أشرف المرسلين سيدنا محمد وعلى آله وصحبه أجمعين.

يُعد علم الإحصاء من العلوم التي شهدت تطورا وارتقاءا واسعا وسريعا في العقود الأخيرة، نظرا لما تلاقيه المقاييس والأساليب الإحصائية من رواج في استخداماتها المتعددة في معظم المجالات والتخصصات العلمية، وأيضا لما يتمتع به هذا العلم من تسارع ومرونة تمكنه من تطوير طرق جديدة ومبتكرة للتعامل مع البيانات المختلفة التي تُنتجها العلوم الحديثة الأخرى باستمرار مهما كان حجم تلك البيانات أو طبيعتها.

ويمكننا القول أن استخدام "الأدوات" الإحصائية الاستكشافية في التعامل مع البيانات هو في الواقع المحور الذي يدور حوله هذا الكتاب بصورة عامة، والهدف هو تمكين القارئ، طالبا كان أم باحثا، من التعرف على أنواع البيانات الأكثر انتشارا من حولنا واختيار الأسلوب أو الأساليب المناسبة للتطبيق بغية استكشاف ما تمثله تلك البيانات أو ما "تخفيه" بين طياتها، ثم اختيار الوسيلة الأفضل لعرض وتفسير المعلومات التي تم التوصل إليها وذلك من خلال الجداول، الرسومات بيانية، أو المخططات الانسيابية.

عن ماذا يتكلم هذا الكتاب؟

يتناول الكتاب المقاييس الإحصائية التي تتعلق بوصف البيانات، وهو ما يعرف حديثا **بالتحليل الإحصائي الاستكشافي** ((Exploratory Data Analysis (EDA))، دون التطرق لتكوين استدلالات أو تقدير لمقاييس المجتمعات، والذي يندرج تحت مفهوم الإحصاء الاستدلالي. إلا أن القسم الثاني من الكتاب يتناول موضوع الاحتمالات وما يتعلق بها من توزيعات واستخدامات وخصائص، وهذا بحد ذاته يُعد المُتطلب الرئيسي الأول لدراسة الإحصاء الاستدلالي.

من المهتمين بهذا الكتاب؟

الكتاب لا يخاطب المتخصصين في علم الإحصاء فقط وإنما يتميز بأسلوبه المُبسط الذي سيجده غير المتخصصين في متناولهم، ويمكن اعتباره مدخل، في طريقة تدرجه وتصنيفه للموضوعات، ومرجعا لدارسي المقررات الإحصائية مثل الإحصاء العام، الإحصاء الحيوي، الاحتمالات، الإحصاء الرياضي، الطرق الإحصائية، الإحصاء التطبيقي، وغيرها مما يدور في هذا المجال.

ماذا يحوي هذا الكتاب من مواضيع إحصائية؟

في الفصل الأول يتم تناول تعريف علم الإحصاء وأنواع البيانات، وكذلك طرق جمعها وتنظيمها في قواعد البيانات البسيطة. ويأتي التحليل الاستكشافي للبيانات ليحتل الفصلين الثاني والثالث، حيث يضم الفصل الثاني الجزء الأول من التحليل الاستكشافي الذي يهتم بتوزيع البيانات وإنشاء الجداول التكرارية، وتمثيل تلك البيانات بيانيا من خلال مجموعة كبيرة من الرسوم والمخططات البيانية (مثل الرسم النقطي، المدرج التكراري، الأعمدة البيانية، ...)، ثم يعرض هذا الفصل أهم مقاييس النزعة المركزية (مثل الوسط، الوسيط، التجزيئات، وغيرها). ويتم استكمال الجزء الثاني من التحليل الاستكشافي في الفصل الثالث الذي يعرض مقاييس التشتت (مثل المدى، الانحراف المعياري، ...)، وأيضا الدرجات المعيارية، العزوم، والرسومات البيانية المتعلقة بتوزيع البيانات مثل (شكل الصندوق، وشكل الساق والورقة). في الفصل الرابع يتم تناول موضوع الاحتمالات وما يحتويه من مفاهيم (مثل الأحداث وفراغ العينة، طرق العد، مسلمات الاحتمال، ...).

وتُستكمل موضوعات الاحتمال في الفصل الخامس الذي يضم تعريف المتغيرات العشوائية والتوزيعات الاحتمالية وخواصها والدوال المتعلقة بها، وذلك في الحالة المفردة والثنائية للمتغيرات. يأتي الفصل السادس مُكملا لموضوع التوزيعات الاحتمالية ويمثل مرجعا في نفس الوقت لأهم التوزيعات الاحتمالية المنفصلة والمتصلة في علم الاحتمالات وخواص تلك التوزيعات. أما الفصل السابع فيتناول موضوع توزيعات المعاينة التي لا غنى عنها في دراسة الإحصاء الاستدلالي ونظرياته.

ويجد القارئ في نهاية كل فصل مجموعة من التمارين العملية والمتعلقة بموضوعات الفصول، وحلول تلك التمارين ستكون متوفرة في الملحق 2 في نهاية الكتاب.

تقدير واعتزاز

أود في هذا الموضع أن أسجل تقديري واعتزازي بعلم الإحصاء، ليس تحيزا أو تعصبا، ولكن تفكرا في هذا العلم الذي يكفيه فخرا أنه الأكثر ذكرا في القرآن الكريم وذلك في أكثر من سورة؛ (... مال هذا الكتاب لا يغادر صغيرة ولا كبيرة إلا **أحصاها** ...، الكهف 49)، (لقد **أحصاهم** وعدهم عدا، مريم 94)، (... **أحصاه** الله ونسوه ...، المجادلة 6)، وغيرها من الآيات.

شُكر

أود أن أشكر كل أساتذتي وزملائي الأفاضل أعضاء هيئة التدريس في قسم الإحصاء بكلية العلوم في جامعة بنغازي على ما تعلمته منهم، وأشكر طلبتي الأعزاء داخل القسم وخارجه، واللذين كما علّمتهم ... تعلّمت معهم.

المؤلف

المحتويات

الفصل الرابع

الفصل الخامس

الفصل السابع

توزيعات المعاينة
(Sampling Distributions)

الفصل الأول

مقدمة: علم الإحصاء والبيانات

(Introduction: Statistics and Data)

1.1 تعريف علم الإحصاء وأهميته (Definition and Importance of Statistics)

2.1 تعريف البيانات في علم الإحصاء (Definition of Data in Statistics)

3.1 أنواع البيانات (Data Types)

4.1 مصادر وطرق جمع البيانات (Sources and Methods of Data Collection)

1.4.1 نظم جمع البيانات (Data Collection Systems)

2.4.1 طرق جمع البيانات من مصادرها (Methods of Data Collection)

3.4.1 أساليب سحب العينات (Sampling Techniques)

5.1 استخدام قواعد البيانات في البحث الإحصائي (Using Databases in Statistical Research)

6.1 تمارين الفصل الأول

1.1 تعريف علم الإحصاء وأهميته (Definition and Importance of Statistics)

إن المجتمعات التي نعيش فيها اليوم تنتشر المعلومات فيها انتشارا كبيرا. هذه المعلومات التي غالبا ما يتم التعبير عنها بالأرقام التي نراها ظاهرة في محيطنا اليومي في كل مكان؛ في التلفاز والمذياع والجرائد والمجلات وعلى شبكة الإنترنت وغيرها من وسائل الإعلام الحديث. وظهور هذه المعلومات قد يكون من خلال جداول وملخصات توضح نسب انتشار ظاهرة اجتماعية مثل ظاهرة التدخين بين الأطفال في إحدى المدن، أو رسومات بيانية مكونة من أعمدة أو دوائر تظهر تفوق قطاع أو خط إنتاج معين في أحد المصانع في سنة ما، أو حتى تقارير إحصائية لمنظمة الصحة العالمية، وغيرها.

في الحقيقة، لسنا بحاجة لبذل جهد كبير للبحث عن وجود علم الإحصاء في حياتنا اليومية، ولكن الفكرة تكمن في كيفية استخدام الأسلوب الصحيح في العمل أو الدراسة وفهم وتحديد معنى المؤشرات الإحصائية وبناء القرار المناسب بالاعتماد عليها.

من ناحية أخرى، فإن كثرة هذه الأرقام والمعلومات قد تصيبنا أحيانا بالإرباك والحيرة، فنشرع في البحث عن حلول لاختزال هذه الأرقام بصورة تمكننا من فهم الصورة بشكل واضح عن طريق استخدام "جزء" فقط من هذه الأرقام. لذلك فإن بعض النتائج الإحصائية الهامة التي نصادفها قد تكون في الحقيقة الناتج النهائي لجهد كبير بدأ بجمع البيانات (والتي سيتم توضيح مدلولها في الجزء القادم) من مصدرها الأصلي ومر عبر سلسلة طويلة تم فيها معالجة وتنظيم وتلخيص هذه البيانات ثم تحليلها وإعداد نتائجها للنشر بصورة واضحة مفهومة. وهنا لابد من الإشارة إلى أن دقة وصحة النتائج الإحصائية يعتمد بشكل كبير على صحة البيانات المستخدمة في التحليل الإحصائي، ويقصد بصحة البيانات هنا مصداقية مصدر البيانات ودقة الجمع والرصد وتجنب الأخطاء. إضافة إلى ذلك، فإن دقة النتائج الإحصائية قد يعتمد أيضا على طريقة جمع تلك البيانات والأدوات الإحصائية المستخدمة في المعالجة والتحليل. فنوع البيانات وطبيعة الدراسة وأهدافها تفرض طريقة التحليل المناسبة والتي ستؤدي بدورها للحصول على المعلومات. وهنا يطرح السؤال التالي نفسه؛ من الذي يحتاج لهذه المعلومات والنتائج الإحصائية؟.

إن الإجابة قد تختلف باختلاف مدى استخدام علم الإحصاء من دولة لأخرى أو حتى من مجتمع لآخر، ولكن بصورة عامة يمكننا القول بأن الشرائح التالية هي الأكثر احتياجا للمعلومات التي يوفرها علم الإحصاء:

الحكومات: فهي تحتاج لتلك المعلومات الإحصائية المتعلقة بالتعدادات والدراسات الاقتصادية والاجتماعية والتعليمية والصحية لغرض مراقبة ما يدور ويطرأ على التركيبة السكانية ووضع الخطط المستقبلية للبناء والتطوير وتحسين الظروف المحيطة بالسكان.

المؤسسات الاقتصادية: فدراسة حركة الأسواق المحلية والعالمية يعد أساسا لإقامة المشاريع الاقتصادية الصغيرة والكبيرة وهذا عادة ما يعتمد على المؤشرات الإحصائية وخاصة المتعلقة بالزمن.

الأفراد: وتضم هذه الشريحة كل من يستخدم الإحصاء سواء في الدراسة أو العمل أو البحث العلمي وتشمل الطلبة في مختلف مراحل التعليم، والإداريون، ومدراء الأقسام التنفيذية، وحتى عامة الناس بغرض مواكبة الأحداث من حولهم.

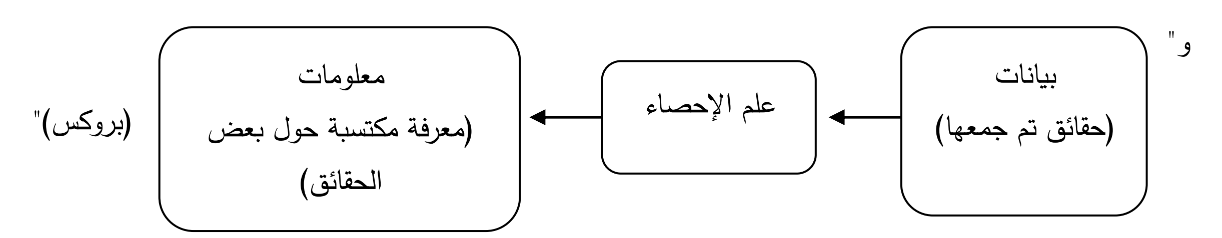
ننتقل الآن لمناقشة أحد أهم المفاهيم التي يرتكز عليها علم الإحصاء ألا وهو مفهوم **التغير** (Variability)، والذي يقع ضمن المفردات أو المشاهدات الخاصة بتجربة أو عملية أو ظاهرة معينة عندما لا تتوالى بشكل ثابت أو لا تعطي نفس النتائج بالضبط، مثل التغير المشاهد في أسعار النفط مثلا خلال عدة سنوات أو تغير درجات الحرارة من منطقة جغرافية لأخرى، أو التغير في مستوى الذكاء ضمن طلبة الفصل الواحد. وهذا في الواقع ما نصادفه ونتعامل معه في حياتنا اليومية، فمعظم ما نراه من حولنا هي عبارة عن أحداث متغيرة لا تثبت على حال واحدة.

وهنا يزودنا التفكير الإحصائي بطرق وأدوات مفيدة لاستخدام وإدراج هذا التغير في فهم طبيعة التجربة أو الظاهرة وكذلك في عملية اتخاذ القرار عند الحاجة. فعلى سبيل المثال، إذا ما تناولنا عملية استهلاك الوقود في سيارة، فهل تعتقد أنك ستقطع نفس المسافة بالضبط عند ملء خزان الوقود في كل مرة؟ الإجابة ستكون لا بالطبع، لأن هذا "التغير" الملاحظ في استهلاك الوقود يعتمد في الواقع على عدة عوامل مثل طريقة قيادة السائق، طبيعة الطرق، العمر الافتراضي لمحرك السيارة ومقدار قوته، وغيرها من العوامل. والأساليب الإحصائية تزودنا بالهيكلية المناسبة لوصف هذه التغيرات ومعرفة مصادرها من خلال التحليل الإحصائي وكذلك تمييز أقواها تأثيرا. وهكذا، يمكننا أن نعرف **المتغيرات** (Variables)، بصورة عامة، بأنها تلك الأشياء التي نقيسها ونراقبها ونتعامل معها من خلال مواجهتنا لظواهر الحياة أو ممارسة التطبيقات والتجارب العملية المختلفة. ومن البديهي أن تختلف هذه المتغيرات في طبيعتها وطريقة قياسها باختلاف المصدر الذي نتجت عنه، وسيتم التحدث عن هذا الاختلاف وما ينشأ عنه من تنوع في أنواع البيانات في البند (3.1).

من هذا المنطلق، نجد أن التفكير الإحصائي يهيمن على معظم المجالات والحقول العلمية، مما دفع ببعض الإحصائيين المعاصرين إلى تعريف علم الإحصاء، (بصورة فلسفية أكثر منها أكاديمية)، على النحو التالي:

"الإحصاء هو طريقة للحصول على المعلومات من البيانات. (كيلر (Keller))"، و "الإحصاء هو أداة لإنشاء مفهوم جديد من مجموعة من المفردات. (بروكس (Brooks))"،

و "



(بروكس)"

من خلال النقاش السابق، يمكننا الآن تعريف علم الإحصاء بأنه ذلك العلم الذي يستخدم في التعامل مع البيانات من خلال ثلاث مراحل أساسية، يتم في الأولى جمع وتنظيم وإعداد قاعدة البيانات الإحصائية، ويتم في الثانية استكشاف ووصف البيانات و/أو تحليلها (إذا ما تطلبت الحاجة)، ويتم في المرحلة الثالثة استخلاص النتائج وعرض المعلومات، (إما لغرض وصف الظاهرة أو لاتخاذ القرار)، وتشمل هذه المرحلة أيضا استخدام المعلومات لغرض التنبؤ ووضع الخطط المستقبلية.

وفي سياق هذا التعريف يمكن تقسيم علم الإحصاء، اعتمادا على الأهداف المطلوبة والطرق المستخدمة في الوصف والتحليل، إلى فرعين رئيسيين هما الإحصاء **الوصفي** أو **الاستكشافي** (Descriptive or Exploratory Statistics) والإحصاء **الاستدلالي** أو **الاستنتاجي** (Inferential Statistics). بالنسبة للفرع الأول فإنه يضم تلك الطرق التي تهتم بوصف مجموعة من البيانات واستكشافها بغرض الحصول على معلومات توضح طبيعة الظاهرة أو السلوك السائد لمشاهدات الدراسة. وأما الإحصاء الاستدلالي فإنه يهتم بتحليل جزء من البيانات لغرض الحصول على دلالات أو

استنتاجات حول مجموعة البيانات الكلية. وحديثا، قد نقرأ في كثير من كتب علم الإحصاء عن مصطلح **التعلم الإحصائي** (Statistical Learning)، حيث يتم إطلاق العنان للبيانات " لتعبر" عن نفسها وتظهر ما تحتويه من معلومات من خلال استخدام عدة أساليب إحصائية لتحقيق هدف واحد وعدم الاكتفاء بأسلوب واحد.

أما من الناحية التطبيقية، فإن الطرق المختلفة لعلم الإحصاء يتم استخدامها في كافة مجالات وميادين الحياة تقريبا، فعلم السكان والمسوح يستخدم مثلا لرسم الخريطة السكانية للدولة أو مناطق معينة داخلها ودراسة حركة الهجرة والولادة والوفاة وغيرها، وكذلك لاستطلاع الرأي العام حول ظاهرة أو قرار سياسي. وأساليب المعاينة تشكل أداة فعالة لتوفير البيانات لدراسة سلوك شريحة من المستهلكين تجاه منتج جديد في السوق مثلا. والتجارب الإحصائية المراقبة تكون مفيدة للأطباء لدراسة تأثير العقاقير المختلفة على الناس. والمهندسون من ناحية أخرى يعتمدون بشكل كبير على أساليب مراقبة الجودة في المصانع والمؤسسات الإنتاجية. والمهتمون بالاقتصاد عادة ما يهتمون بدراسة المؤشرات الإحصائية المختلفة خلال فترات زمنية معينة لتحديد نمط اقتصادي سائد أو للتنبؤ بالحالة الاقتصادية المحلية أو العالمية في السنوات المقبلة.

وهكذا فإننا نرى أن الأساليب والأدوات الإحصائية تلعب دورا هاما في تحقيق الهدف المنشود في الأمثلة السابقة. وقد ذهب بعض الباحثون إلى القول بأن علم الإحصاء هو في الواقع علم "خدمي"، أي علم يخدم العلوم الأخرى، وفي هذا القول جزء من الحقيقة، إلا أنه يجب الإقرار بأن علم الإحصاء هو علم قائم بذاته، (رغم اعتماده بصورة كبيرة على المفاهيم والقواعد الرياضية من عمليات أساسية وتفاضل وتكامل ومعادلات خطية وغير خطية وهندسة فراغية وغيرها)، إلا أن هذا الاعتماد لا يفقده استقلاليته وتميزه.

وبشكل عام، يمكن القول أن الأسلوب والمنهج الإحصائي في التعلم يزودنا بالقواعد والطرق التي يمكن استخدامها لاستكشاف ووصف الظواهر المختلفة في الحياة، وكذلك التوصل للقرارات الصحيحة (بنسب عالية من الدقة المنطقية) باستخدام الأساليب الإحصائية التي ظهرت في العقد الأخير. إضافة إلى أن التخطيط العلمي السليم، سواء كان اقتصاديا أو اجتماعيا أو طبيا أو غير ذلك، يجب أن يسانده ويدعمه الرأي الإحصائي من خلال استخدام المنهج الإحصائي المناسب واعتمادا على قاعدة بيانات صحيحة ودقيقة.

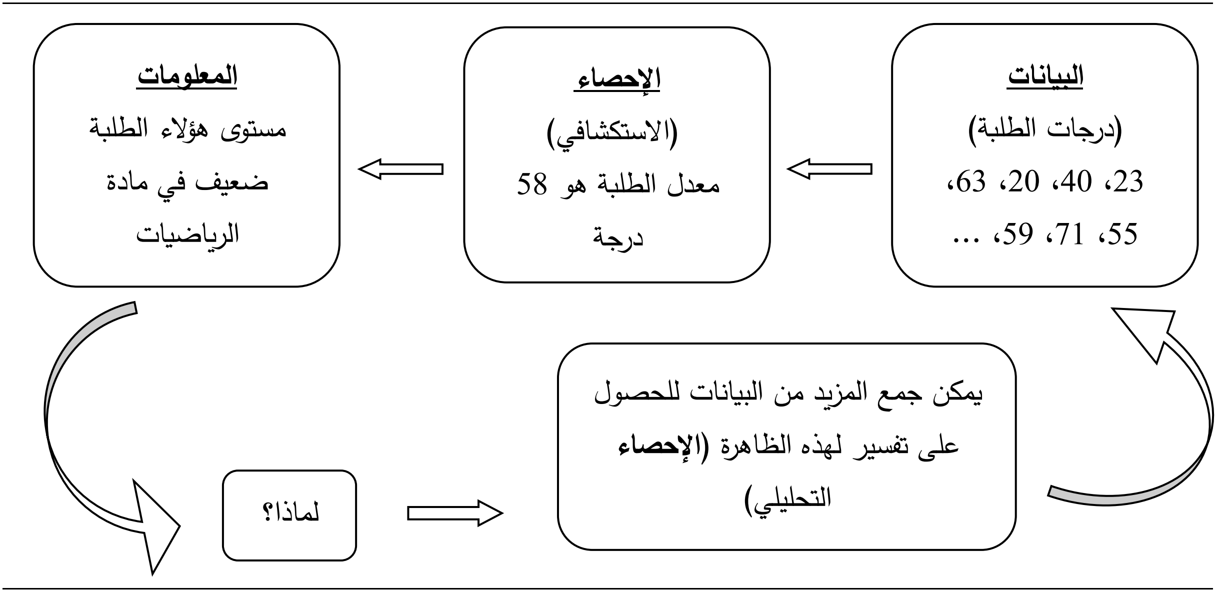
2.1 تعريف البيانات في علم الإحصاء (Definition of Data in Statistics)

قبل أن نتمكن من الحصول على المعلومات المتعلقة بدراسة معينة بغية تحليلها واتخاذ القرار المناسب أو حتى لغرض العرض والاستكشاف فقط، يجب أولا الحصول على **البيانات** (Data). فكما أن الأشجار هي المادة الخام التي نحصل من خلال تصنيعها على مادة الورق، فالبيانات بالمثل تمثل المادة الخام التي من خلال التعامل معها نحصل على المعلومات المفيدة. ولفظ "بيانات" يظهر كثيرا في الكتب والمطبوعات ووسائل الإعلام بل وحتى في التعاملات اليومية. وهنالك في الحقيقة عدة تعريفات لمصطلح "البيانات"، ضمن إطار علم الإحصاء، نسوق منها على سبيل المثال ما يلي:

"البيانات هي الحقائق أو الأرقام التي يتم الحصول على النتائج من خلالها. (الهيئة الكندية لإحصاء)"، و"البيانات هي مجموعة من المشاهدات و/أو المقاييس المأخوذة من جزء من المجتمع للحصول على معلومات معينة أو الإجابة على سؤال محدد. (لا بلانك (LeBlanc))".

في الحقيقة، إن مصطلح بيانات هو مصطلح شمولي، فالبيانات الشخصية التي تضم الاسم، العمر، النوع، الحالة الاجتماعية، وغيرها هي نوع من البيانات، والتاريخ الطبي للمريض والذي يضم قراءات نسب السكر، ضغط الدم، درجات الحرارة، وغيرها هي بيانات، وكذلك المشاهدات الناتجة عن تجربة كيميائية لقياس تفاعل مركبين هي أيضا بيانات، وهكذا. لذلك يمكننا القول بأن أي ظاهرة أو دراسة أو تجربة أو حتى مراقبة لعملية معينة يمكن أن ينتج عنها جميعا بيانات. فالبيانات هي المقياس الفعلي الذي نحصل عليه في النهاية.

وأحيانا يتم الخلط بين مفهوم "البيانات" ومفهوم "المعلومات"، وهما في الواقع مختلفان، والشكل (1.1) يعطي مثلا بسيطا يمكن من خلاله توضيح المفهومين وعلاقة علم الإحصاء بهما.



شكل (1.1): هيكلية تعامل علم الإحصاء مع البيانات والمعلومات من خلال بيانات خاصة بدرجات مجموعة من الطلبة في مادة الرياضيات في إحدى الكليات العلمية، (الدرجة من 100).

أي أن الأساليب الإحصائية تقوم بالتعامل مع البيانات بصورة استكشافية أو تحليلية بغية الحصول على المعلومات المطلوبة بلغة الأرقام.

إن قواعد البيانات البسيطة، (والتي سنتناولها بشيء من التفصيل لاحقا في هذا الفصل)، عادة ما يكون لها بناء محدد مكون من **مفردات** (Individuals) أو **مشاهدات** (Observations) ناتجة عن مقاييس لمتغير أو أكثر. والجدول (1.1) يوضح بناء قاعدة بيانات تقليدية بسيطة. والمصطلح "متغير" يشير، في علم الإحصاء، إلى الصفة المميزة (Characteristic) لشيء أو مجموعة من الأشياء والتي يمكن التعبير عنها بقيمة (رقم) بحيث يأخذ المتغير أكثر من قيمة. فأي ظاهرة يمكن قياسها بوحدة قياس مناسبة يمكن أن يعبر عنها بمتغير، فالطول (بالسنتيمتر أو القدم) لمجموعة من الأشخاص هو متغير، والعمر الاستهلاكي (بالأشهر أو السنوات) لبطاريات السيارات هو متغير، وتقديرات طلبة السنة الثانية في كلية ما (ممتاز، جيد جدا، ...) تمثل أيضا متغير. وسيتم توضيح أنواع المتغيرات في البند القادم، (3.1).

في الجدول (1.1) يُلاحظ أن كل صف يُعرف حالة واحدة لكل المشاهدات الموجودة في المتغيرات لمفردة واحدة، وكل عامود يحتوي على كل المشاهدات المعرفة لمتغير واحد. وعادة فإن المعلومات التي يتم استخلاصها من قواعد البيانات

تكون موجودة ضمن المتغيرات والمشاهدات بطريقة غير ظاهرة بصورة مباشرة، فلا يمكننا بمجرد النظر المباشر أن نحصل على ما نريده من معلومات حول ما تصفه البيانات. لذلك نستخدم الأساليب الإحصائية المختلفة لسبر أغوار هذه البيانات واستخراج ما تحويه من معلومات مفيدة (إن وجدت في الأصل).

جدول (1.1): قاعدة بيانات خاصة بدراسة افتراضية.

توصيف الظاهرة وما تمثله المتغيرات

المشاهدة	المتغيرات			
	اسم المتغير الأول (X_1)	اسم المتغير الثاني (X_2)	…	اسم المتغير الأخير (X_p)
1	122	خفيف	…	3.44
2	412	ثقيل جدا	…	5.09
3	312	ثقيل	…	1.68
⋮	⋮	⋮	⋮	⋮
n	177	متوسط	…	7.14

وهذا ما يؤكده أيضا بعض الإحصائيين، "إن الخطوة الأولى لفهم البيانات هي أن نستمع لما تريد أن تقوله لنا، وأن ندع الإحصاء ليتكلم عن نفسه. (مور (Moore))". وبما أن البيانات لن تتكلم بمفردها، فإننا نقوم "بمساعدتها" عن طريق إتباع الإجراءات الإحصائية من تنظيم وتلخيص وتحليل وعرض بصورة مناسبة للوصول إلى المعلومات الكامنة، أو القرار المناسب.

3.1 أنواع البيانات (Data Types)

ذكرنا في البند السابق أن البيانات هي المادة الخام للمعلومات، ومن أجل الوصول للمعلومة الصحيحة واتخاذ القرار السليم، لابد من ضمان استخدام الوسيلة أو الأسلوب الإحصائي (الاستكشافي أو التحليلي) المناسب. والأساليب الإحصائية تعتمد بدورها من حيث التطبيق والعرض على طبيعة ونوع البيانات، وفي هذا الجزء سنلقي الضوء على أنواع البيانات التي يتم التعامل معها في علم الإحصاء.

يمكن من حيث الطبيعة أن نقسم أنواع البيانات إلى نوعين رئيسيين؛ البيانات **الكمية** (Quantitative Data) والبيانات **النوعية** أو **الوصفية** (Qualitative Data)، أو قد تسمى أيضا بالبيانات **القطاعية** أو **التصنيفية** (Categorical Data).

البيانات الكمية تحتوي مشاهدات تم قياسها ورصدها مباشرة بصورة أرقام لها قيمة كمية، مثل درجات الحرارة المسجلة يوميا في مدينة مكة المكرمة، أو أوزان الأطفال حديثي الولادة في أحد الشهور. فهذه المشاهدات يتم التعبير عنها بنقاط (Scores) مأخوذة على معيار أو مقياس محدد. وهذا النوع من البيانات تمثله متغيرات يمكن التعامل معها بواسطة العمليات الحسابية مباشرة.

من ناحية أخرى، توجد بعض الظواهر والتجارب التي يتم التعبير عنها بصفات لا بأرقام، مثل النوع أو الجنس (ذكر، أنثى)، أو تقسيمات فصائل الدم (A، B، AB، O)، فهذه البيانات تسمى بيانات نوعية أو قطاعية حيث أنها تأخذ قيما تحدد تصنيف المفردة إلى نوع أو فئة أو قطاع محدد ولا يمكن التعامل معها بصورة مباشرة[1] بالطرق الحسابية التقليدية، إلا أنه يمكن استخدام رموز (Codes) رقمية عوضا عنها لتسهيل التعامل معها بصورة إحصائية.

إن طبيعة المعلومات التي يمكن أن نحصل عليها من البيانات يتحكم بها عامل مهم هو **نوع المقياس** (Measurement Scale) المستخدم في رصد البيانات. من هذا المنطلق يمكن تقسيم البيانات الكمية إلى نوعين أساسين هما:

1) بيانات **المقياس الفئوي** (Interval Scale): وهذا النوع يسمح لنا بترتيب المفردات على مقياس محدد بحيث يمكن تحديد القيمة الفعلية لكل مفردة. وكذلك يمكن المقارنة باستخدام الفرق بين هذه المفردات، إلا أن المقياس الفئوي لا يقيس الصفر كقيمة. ومن الأمثلة على هذا النوع من البيانات قياس درجة الحرارة (بالقياس المئوي أو الفهرنهيت)، فهذا القياس له ترتيب، فيقال مثلا أن درجة الحرارة 50° هي أكثر من 40°، والزيادة من 20° إلى 40° مثلا هو ضعف الزيادة من 30° إلى 40°. وهذا المقياس يمثل قيمة مرجعية مطلقة غير نسبية، أي أن القيم فيه قد تتراوح في العموم ضمن الفترة الحقيقية (∞، $-\infty$) ويكون للفرق بين أي قيمتين معنى، إلا أن النسبة بين قيمتين لا يكون لها معنى، فمثلا لا معنى لأن نقول أن درجة الحرارة اليوم هي بنسبة 1.5 ضعف درجة حرارة الأمس، وهكذا.

2) بيانات **المقياس النسبي** (Ratio Scale): وهي تشبه بيانات المقياس الفئوي إضافة إلى أنها تعتبر الصفر من ضمن درجات القياس، لذلك فإنها تسمح بحساب النسبة بين قيمتين ضمن المفردات، مثل أن نقول أن ولدا عمره 10 سنوات هو أكبر بمرتين من ولد عمره 5 سنوات. وهذا النوع من المقاييس يكون شاملا للعديد من أنواع البيانات الكمية مثل العمر والوزن والطول والزمن والمبالغ النقدية وغيرها.

أما البيانات النوعية فتقسم هي الأخرى، من حيث المقياس، إلى قسمين هما:

1) بيانات **المقياس الاسمي** (Nominal Scale): وهي تعبر عن تصنيف المفردات إلى فئات أو قطاعات مختلفة، إلا أننا لا نستطيع تحديد القيمة الفعلية لكل فئة، ولا حتى ترتيب هذه الفئات بشكل تصاعدي أو تنازلي. فمثلا، لا يمكن القول أن الذكور أفضل من الإناث أو العكس، أو القول بأن لون معين هو أحسن من الآخر. ومن أمثلة هذا النوع من البيانات الإجابات المتضادة (نعم، لا)، وتحديد المدن (بنغازي، درنة، طرابلس، ...)، والألوان (أبيض، أزرق، أحمر، ...)، وغيرها.

---

[1] يجب التنويه هنا إلى أن الكثير من البرامج الإحصائية الحديثة تتيح للمستخدم إمكانية التعامل مع البيانات الغير رقمية بصورة مباشرة دون الحاجة لإعطائها رموزا رقمية.

2) بيانات **المقياس الترتيبي** (Ordinal or Rank Scale): على عكس نوع المقياس السابق، فإن طبيعة هذه البيانات تسمح بترتيب المفردات وفق نظام معين من الأقل قيمة إلى الأكثر قيمة، (كما هو المعتاد)، إلا أنه لا يمكن حساب الفرق في القيمة بين أي فئتين من الفئات المرتبة. فمثلا في بعض الدراسات الاجتماعية- الاقتصادية يمكن تقسيم الدخل إلى الفئات (تحت المتوسط، متوسط، فوق المتوسط)، إلا أننا لا نستطيع القول، مثلا، بأن فئة فوق المتوسط هي أعلى بـ 20% من فئة متوسط. والأمثلة على هذه النوعية من البيانات هي كثيرة، فالحجم، مثلا، يمكن تقسيمه إلى (صغير، متوسط، كبير، كبير جدا)، درجة اللون يمكن تقسيمها إلى (فاتح، معتدل، غامق)، وحتى آراء مجموعة من الأطباء حول إجراء عملية جراحية يمكن تقسيمها إلى (معارض بشدة، معارض، حيادي، مؤيد، مؤيد بشدة)، وهكذا. ولاحظ أنه، كما ذكرنا سابقا، لا معنى لمفهوم أفضل أو أسوء للقيم في هذه التقسيمات.

وتجدر الإشارة أيضا إلى أن المتغيرات التي تضمها البيانات يمكن أيضا تصنيفها[1] إلى متغيرات **مستمرة** أو **متصلة** (Continuous) ومتغيرات **متقطعة** أو **منفصلة** (Discrete).

فالمتغيرات التي يتم قياس مفرداتها من خلال قيم حقيقية (تشمل القيم العشرية أيضا) ضمن فترة محددة، (وإن كانت مستقلة عن بعضها البعض)، تسمى متغيرات متصلة، ومثال على ذلك الوزن بالكيلوجرام أو الرطل، ونسبة الحموضة (ph)، وتركيز الهيموجلوبين في الدم، وغيرها. أما المتغيرات الناتجة عن حصر نتائج الظاهرة بالعد (Count)، مثل عدد مرضى ضغط الدم المسجلين في أحد المستشفيات، أو عدد المساجد في مدينة اسطنبول، أو تقديرات الطلبة في مقرر دراسي (A, B, C, D, F)، فجميعها تكون متغيرات منفصلة أو متقطعة.

4.1 مصادر وطرق جمع البيانات (Sources and Methods of Data Collecting)

إن طرق جمع البيانات تعتمد في الواقع على عدة عوامل أهمها طبيعة هذه المصادر وكيفية وجود البيانات فيها، وكذلك طبيعة وأهداف الدراسة المطلوب توفير البيانات لها. لذلك كان من الضروري مناقشة كلا من الموضوعين؛ مصادر البيانات، وطرق جمع البيانات معا في هذا الجزء. وسنتناول أولا نظم جمع البيانات العامة، ثم نتطرق للطرق التفصيلية لعملية الجمع من المصادر الرئيسية، وبعد ذلك نستعرض أهم طرق المعاينة الإحصائية.

1.4.1 نظم جمع البيانات (Data Collection Systems)

توجد ثلاث نظم رئيسة لجمع البيانات في الحياة العملية، وهي التعداد، المسح العيني، والبيانات الإدارية. ويعتمد استخدام أي نظام من هذه الأنظمة على عدة عوامل منها طبيعة الدراسة أو البحث وأهدافه، ودرجة الدقة المطلوبة في البحث، ومدى المصداقية المتوفرة في مصدر البيانات وغير ذلك من العوامل. وسنقوم باستعراض كل نظام من هذه النظم على حده موضحين مزاياه وعيوبه.

1) **التعداد** (Census): في هذا النظام يتم جمع البيانات حول كل المشاهدات المتوفرة من المصدر. فمثلا إذا اعتبرنا، تجاوزا، أن جامعة بنغازي هي المصدر الذي يحتوي على "كل" البيانات، فإننا نستطيع جمع بيانات

---

[1] سيتم تعريف أنواع المتغيرات بصورة أشمل عند تناول مفهوم المتغيرات العشوائية.

تتعلق بالأعمار أو المعدلات الدراسية أو الحالة الاجتماعية والاقتصادية لكل الطلبة، ونكون بذلك قد استخدمنا طريقة التعداد في الجمع.

ومن مميزات نظام التعداد أن الأخطاء الناتجة عن الجمع أو التحليل الإحصائي المطلوب إجراؤه لاحقا تكون محدودة، وكذلك يتم عادة الحصول على أدق التفاصيل المتعلقة بالمشاهدات الموجودة في مصدر البيانات وهذا يزودنا بإدراك أوسع وأشمل خلال مرحلة البحث. أما عيوب نظام التعداد فتتلخص في كونه باهظ التكلفة ويستغرق جهدا ووقتا طويلا عادة في تنفيذها، إضافة إلى ظهور ما يعرف **بعبء الاستجابة** (Response Burden) حيث يتطلب نظام التعداد تسجيل بيانات حول كل مفردة من مفردات المصدر على حده مما يشكل عبئا كبيرا على الباحثين وجامعي البيانات، إضافة إلى صعوبة إبقاء كل المفردات تحت المراقبة والتحكم المستمر خلال فترة التعداد.

قبل أن ننتقل للنظام التالي يجب توضيح مفهومين هامين دائما ما يستخدما خلال عملية جمع البيانات وكذلك في عملية التحليل الإحصائي عموما وهما؛ **المجتمع** و**العينة**. فللقيام بأي دراسة علمية أو لمحاولة حل مشكلة معينة أو حتى للتحقيق في أي موضوع والوصول للإجابة الشافية، يجب تركيز الاهتمام حول مجموعة معينة من المفردات، مثل مجموعة من الناس، المدن، نتائج الامتحانات، تركيز مواد كيميائية، ... ، فالبيانات التي تتكون من كل المفردات التي هي محط اهتمامنا تكون المجتمع، والذي يمكن تعريفه بالصورة التالية:

تعريف (1.1): **المجتمع** (Population): هو مجموعة البيانات التي تتكون من كل المشاهدات محل الدراسة والاهتمام.

من ناحية أخرى لنأخذ المثال التالي؛ عندما نرغب بمعرفة درجة نضوج الطعام فإننا بطبيعة الحال لا نتناوله كله، بل "نتذوق" منه جزءا قليلا (عينة)، ونقوم بإبداء الرأي فيه ثم نقوم "بتعميم" رأينا على الطعام كله (المجتمع). وهذا يعكس لنا مفهوم العينة.

تعريف (2.1): **العينة** (Sample): العينة هي جزء (صغيرا كان أو كبيرا) من المجتمع.

2) **المسح العيني** (Sample Survey): في نظام المسح العيني يتم اختيار جزء فقط من كل البيانات المتوفرة في المصدر. فاختيار وتسجيل بيانات حول 200 طالب مثلا، من جامعة القاهرة يسمى ببساطة مسح عيني حول الطلبة. ومن مميزات هذا النظام تغلبه على عيوب نظام التعداد من حيث تقليص تكلفة عملية جمع البيانات، وإمكانية تنفيذها بوقت أقصر، إضافة إلى السهولة في التحكم والتعامل مع عدد أقل من المفردات.

أما عيوب المسح العيني فهي إمكانية ظهور أخطاء في عملية المسح بسبب التعامل مع جزء فقط (عينة) من المجتمع وإهمال الباقي. فقد لا تكون هذه العينة التي تم سحبها معبرة أو ممثلة لحقيقة المجتمع أو أنها قد لا تحتوي على كل التفاصيل والمعلومات الموجودة في المجتمع[1].

3) **البيانات الإدارية** (Administrative Data): وهي البيانات التي تنتج عن سير العمليات الإدارية المختلفة في الدولة، وتشمل كافة البيانات الموجودة في المؤسسات والهيئات والمراكز الرسمية التي تقدم الخدمات للناس

---

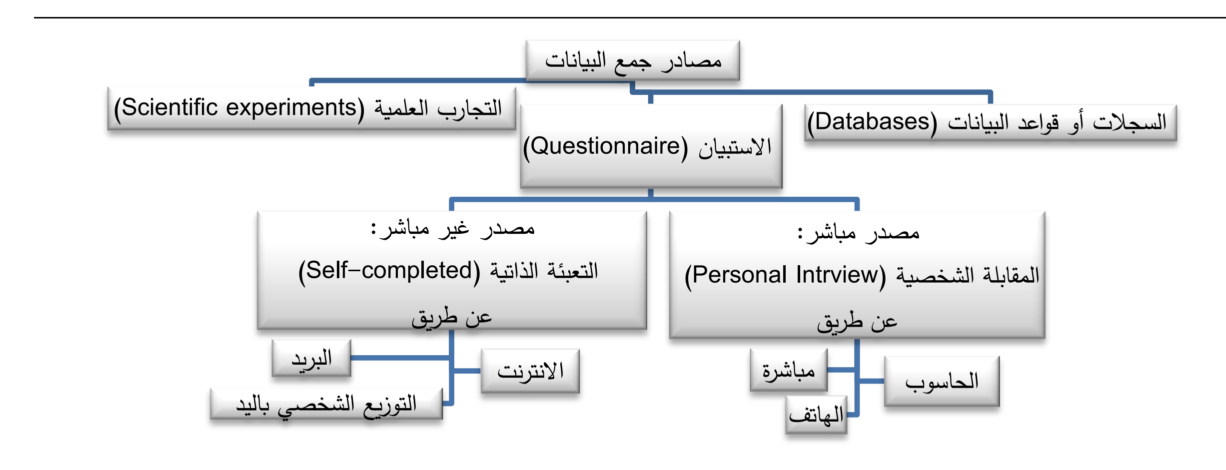
[1] عادة ما يتم تناول هذه النقطة بالتفصيل تحت مفهوم الاستدلال الإحصائي.

بصورة دورية منتظمة، مثل الإدارات الحكومية (تعليم، تخطيط، إسكان، مالية، ...)، وكذلك المراكز الصحية مثل المستشفيات والعيادات، وأيضا المصانع الإنتاجية، وغيرها. فينتج عن ذلك نظام البيانات الإدارية والذي يفترض به أن يكون دقيقا ومنظما، وتكون البيانات فيه موثقة ومحفوظة في ملفات (ورقية أو الكترونية)، وجاهزة الاستخدام الرسمي الإداري (وحتى الإحصائي) بغرض استخلاص وعرض المعلومات حين الحاجة. ويمكن في الواقع تطبيق نظامي التعداد والمسح العيني باستخدام البيانات الإدارية.

وتتميز البيانات الإدارية بقلة الأخطاء لأنها عادة تكون مسجلة بشكل رسمي دقيق، إضافة إلى انخفاض عبء الاستجابة إلى أقل ما يمكن، وأيضا احتواؤها على تسلسل زمني منتظم لعملية التسجيل. أما عيوبها فتتمثل في كونها محدودة أحيانا، فقد تكون هنالك حاجة، على سبيل المثال، لإجراء دراسة طبية متخصصة تتطلب عدة متغيرات (متضمنة لمشاهدات) قد لا يكون بعضها متوفرا في البيانات الإدارية الخاصة للمستشفيات ضمن مركز الدراسة. أضف إلى ذلك أن البيانات الإدارية قد لا تواكب التغير والتطور خلال السنوات المتلاحقة (كما هو الحال في معظم الدول النامية) حيث أن بعض التعريفات العلمية أو المتغيرات قد تستحدث بحيث لا نجد لها مرادفات في بيانات إدارية قديمة. قصور آخر يؤخذ على نظام البيانات الإدارية وهو أن مُعدي منظومات الحاسوب، والتي يتم من خلالها إدخال البيانات، قد لا يكونون من الملمين بالجانب الإحصائي مما يدفعهم لإهمال الكثير من المتغيرات الهامة والتي تستخدم عادة في التحليل الإحصائي. ناهيك عن وجود اختلافات في طريقة تسجيل ورصد البيانات من فرع لآخر لنفس الإدارة أو الهيئة، في حالة عدم وجود شبكة ربط بينها، مما يشكل صعوبة لمستخدمي هذه البيانات في المستقبل.

2.4.1 طرق جمع البيانات من مصادرها (Methods of Data Collection)

سنستعرض الآن أهم الطرق التقليدية والحديثة المتبعة في جمع البيانات وهذا يعتمد، كما أشرنا سابقا، على طبيعة مصدر البيانات. والمخطط في شكل (2.1) يوضح مصادر البيانات الأساسية والانسيابية في طرق جمع البيانات كما هو متبع في الدول التي تسلك النهج الإحصائي الحديث.



شكل (2.1): مخطط يوضح تقسيم مصادر البيانات وطرق جمعها.

ولاحظ أن طرق الجمع في المخطط كلها تنطوي تحت الأنظمة الثلاث المذكورة سابقا، فمثلا طريقة الاستبيان يمكن تطبيقها على المجتمع والعينة على حد السواء، اعتمادا على طبيعة أو هدف البحث. وتجب الإشارة هنا إلى أن مجتمع الدراسة (وهو المجتمع الذي يحتوي على كل المشاهدات الخاصة بالدراسة)، يمكن تقسيمه إلى **مجتمع الهدف** (Target Population) والذي يضم كل المشاهدات محل الاهتمام، و**مجتمع نطاق العينة** (Sample-domain Population) والذي يضم كل المشاهدات التي يمكن سحب العينات منها. ونسوق المثال التالي للتوضيح:

عند إجراء دراسة تتعلق بالمشاكل الاجتماعية والنفسية التي يصادفها طلبة أقسام الإحصاء في ليبيا، فإن مجتمع الهدف في هذه الحالة يكون جميع طلبة الإحصاء في الجامعات الليبية، ومجتمع نطاق العينة هو جميع الجامعات الليبية التي يوجد بها أقسام للإحصاء ويمكن اختيار الطلبة (العينات) منها، مثل جامعة بنغازي، جامعة طرابلس، جامعة عمر المختار، ... وهكذا.

3.4.1 أساليب سحب العينات (Sampling Techniques)

إن سحب العينات من المجتمع يعتمد في الأسلوب المتبع على طبيعة المجتمع وخطة وأهداف الدراسة. وكما تمت الإشارة سابقا، فإن الصعوبات التي تواجه الباحثين عند دراسة جميع مفردات المجتمع تجعلهم يلجئون إلى اختيار عينة أو عينات من المجتمع بأسلوب معين بحيث تكون صورة مصغرة لهذا المجتمع قدر الإمكان. وبالطبع فإن هذه الصورة لن تكون واضحة وذات قيمة علمية إلا إذا تم اختيار مفردات العينة بطريقة تضمن أن تكون ممثلة للمجتمع تمثيلا صادقا. هذه العينات يطلق عليها اصطلاحا اسم **العينات العشوائية**[1] (Random Samples)، وفيما يلي نسرد أهم أساليب سحب العينات العشوائية:

1) **العينة العشوائية البسيطة** (Simple Random Sample): وهي عينة يتم اختيارها بحجم معين، (يرمز له بالرمز n عادة)، بحيث يكون لكل مفردة من مفردات المجتمع، (والذي يرمز لحجمه بالرمز N)، نفس فرصة الاختيار. وهذا الأسلوب هو الأكثر شيوعا، ويستخدم عند التعامل مع المجتمعات المتجانسة (Homogeneous) والتي لا تحتوي على أقسام أو طبقات مختلفة. ويمكن تطبيق هذا النوع من المعاينة باستخدام **جداول الأرقام العشوائية** (Random Numbers Tables) كما هو موضح في المثال التالي:

مثال (1.1): إذا كان المطلوب هو اختيار عينة مكونة من $n = 12$ شخصا من المرضى من كل مستشفيات الأمراض النفسية (المجتمع) في إحدى الدول، والتي تضم $N = 800$ مريض نفسي، فإننا نقوم بما يلي:

أ. نقوم بترقيم جميع عناصر المجتمع بأرقام مسلسلة تبدأ من 001 وتنتهي بالعدد 800.

ب. من جدول الأرقام العشوائية (ملحق 1، جدول م1)، نختار أحد الصفوف عشوائيا ثم نختار أعمدة متجاورة من الأرقام عشوائيا أيضا. ليبدأ اختيارنا من الصف الثاني والأعمدة من 11 إلى 13 نزولا، (لأن $N = 800$ يتكون من ثلاث خانات)، مع إهمال كل الأرقام المتكررة وأيضا التي هي أكبر من 800، وهكذا فإن مفردات العينة (المرضى) ستكون تلك المناظرة للأعداد: 047، 407، 408، 577، 581، 001، ...، 593.

[1] سيتم مناقشة المفهوم الإحصائي للعشوائية عند تناول مفهوم الاحتمالات في الفصل الرابع من الكتاب، ونكتفي هنا بالتعريف اللغوي **للعشوائية** (Randomness) بأنها حدوث الشيء بصورة عرضية غير مفتعلة، (قاموس ميريام وبستر).

2) **العينة الطبقية** (Stratified Sample): يستخدم هذا الأسلوب من المعاينة عندما يكون المجتمع مقسما إلى مجموعات أو طبقات مختلفة بحيث تتشابه عناصر كل طبقة أو تتمتع بصفة معينة. عندئذ يتم تقسيم المجتمع الذي حجمه N إلى $N_1, N_2, \ldots , N_k$ طبقة، ويتم سحب $n_1, n_2, \ldots , n_k$ عينة من كل طبقة على الترتيب فيكون عدد المفردات في كل عينة مسحوبة من كل طبقة محسوبا باستخدام القاعدة التالية؛

$$n_i = \frac{N_i}{N} \times n \ \ , i = 1,2, \ldots , k \qquad (1.1)$$

مثال (2.1): يوجد مجتمع حجمه 1000 عامل يمثل عدد العمال في أحد مصانع الأزياء. وتم التخطيط لإجراء دراسة تتناول كفاءة أداء العمال بالنظر لمستواهم التعليمي، فتم تقسيمهم إلى 4 طبقات كما هو موضح في الجدول المرفق.

المستوى التعليمي				
أمي	ابتدائي	إعدادي	ثانوي فأكثر	المجموع
400	300	200	100	1000

فإذا كان المطلوب سحب عينة عشوائية حجمها 20 عامل من المصنع (المجتمع) فإن حجم العينة في كل طبقة يمكن حسابه باستخدام القاعدة (1.1) كالتالي:

الطبقة 1: أمي $n_1 = 400/1000 \times 20 = 8$

الطبقة 2: ابتدائي $n_2 = 300/1000 \times 20 = 6$

الطبقة 3: إعدادي $n_3 = 200/1000 \times 20 = 4$

الطبقة 4: ثانوي فأكثر $n_4 = 100/1000 \times 20 = 2$

ويكون $2 + 4 + 6 + 8 = 20 = n$. بمعنى أن يتم اختيار 8 عمال من الطبقة الأولى، و 6 عمال من الطبقة الثانية، ...، وهكذا، باستخدام أسلوب العينة العشوائية البسيطة.

3) **العينة العنقودية** (Cluster Sampling): وفي هذا الأسلوب يتم تقسيم المجتمع إلى مجموعات فرعية لا يشترط تجانسها، وهذه المجموعات يتم تقسيمها إلى مجموعات أخرى وهكذا بحيث تسمى أصغر مجموعة فرعية بالعنقود. ثم يتم اختيار عينة عشوائية بسيطة من كل عنقود على أن يتم دمجها في النهاية، وهذا ما يعرف بالمعاينة العنقودية ذات **المرحلة الواحدة** (One-stage Clustering). أما إذا ما تم اختيار مفردات العينة من أجزاء معينة من التقسيم العنقودي الكلي فإن المعاينة تعرف عندها بالمعاينة العنقودية **متعددة المراحل** (Multi-stage Clustering). وعادة ما يتم استخدام أسلوب العينة العنقودية في المعاينة عند التعامل مع المجتمعات الكبيرة جدا أو المجتمعات المختلفة في تركز البيانات فيها.

مثال (3.1): تم طرح تساؤل حول قدرة سوق العمل في ليبيا على استيعاب خريجي المعاهد العليا بمختلف تخصصاتهم. وكان من ضمن خطوات بحث هذا التساؤل سحب عينة عشوائية مناسبة، والتي يفضل أن تكون عنقودية لأن مجتمع الدراسة مصدره خريجي المعاهد والتي تشمل عدة تخصصات مختلفة وكل معهد بدوره يشمل عدة أقسام. في هذه الحالة يتم سحب عينة عشوائية بسيطة من خريجي كل قسم لكل تخصص في هذه المعاهد، ثم يتم تجميعهم في عينة واحدة هي العينة العنقودية.

4) **العينة المنتظمة** (Systematic Sample): ويستخدم هذا الأسلوب في المعاينة عند عدم توفر قوائم محددة لعدد مفردات المجتمع، أو عندما يكون حجم المجتمع غير ثابت، فيتم اختيار مفردة عشوائيا من المجتمع ثم اختيار المفردة الثانية بعد k (عدد معين) مفردة وهكذا حتى الحصول على حجم العينة المطلوب.

مثال (4.1): في دراسة لمعرفة مدى رضى مجموعة من عملاء أحد المصارف التجارية عن الخدمات التي يقدمها هذا المصرف، تم استخدام أسلوب العينة المنتظمة في السحب لأن عدد عملاء المصرف في حالة تغير دائم. فإذا كان حجم العينة المطلوب هو 30 عميلا، عندها يتم اختيار العميل الأول ثم اختيار العميل رقم 5 (إذا كانت $k = 5$ مثلا) ثم العميل رقم 10 وهكذا حتى يتم اختيار $n = 30$ عميلا.

وإضافة للأساليب السابقة، هنالك أسلوب آخر للمعاينة، وإن كان قليل الاستخدام، هو أسلوب المعاينة **الغرضية** أو **القصدية** (Objective or Non-random Sampling) وفيه يتم اختيار العينة بصورة قصدية غير عشوائية وذلك للحصول على معلومة ما بصورة سريعة، أو لبحث مدى واقعية استبيان معين قبل المباشرة بتوزيعه على المستهدفين وهكذا.

ونختم هذا الجزء بتلخيص أهم الحالات الرئيسية التي تستخدم عندها أساليب المعاينة المختلفة بدلا من التعامل مع كل مفردات المجتمع مباشرة:

1) عندما يكون عدد مفردات المجتمع (حجم المجتمع) كبير جدا بحيث يصعب (أو يستحيل أحيانا[1]) جمعها، وتعرف مثل هذه المجتمعات بالمجتمعات **الغير منتهية** (Infinite Populations)، أو حين يستغرق ذلك الكثير من الجهد والوقت والوسائل التي قد لا تتوفر في كثير من الدراسات.

2) الحالات التي يصعب فيها تحديد جميع مفردات المجتمع مثل دراسة أذواق مستهلكي العطور الغربية بهدف تطوير هذه الصناعة محليا (أو عربيا)، فيصعب في هذه الحالة تحديد عدد هؤلاء المستهلكين بدقة.

3) الحالات التي تتطلب إجراء تجارب على مفردات المجتمع أولا بغية تسجيل النتائج كبيانات أولية للدراسة، مثل إجراء دراسة حول العمر الاستهلاكي لنوع من الغسالات الآلية، حيث لا يمكن تشغيل ومراقبة كل إنتاج المصنع[2] (المجتمع)، وكذلك في الأبحاث البيولوجية المتعلقة بتحليل الدم، إذ أنه من غير المنطقي استخدام كمية الدم الكلية (المجتمع) الموجودة في جسم الإنسان.

5.1 استخدام قواعد البيانات في البحث الإحصائي (Using Databases in Statistical Research)

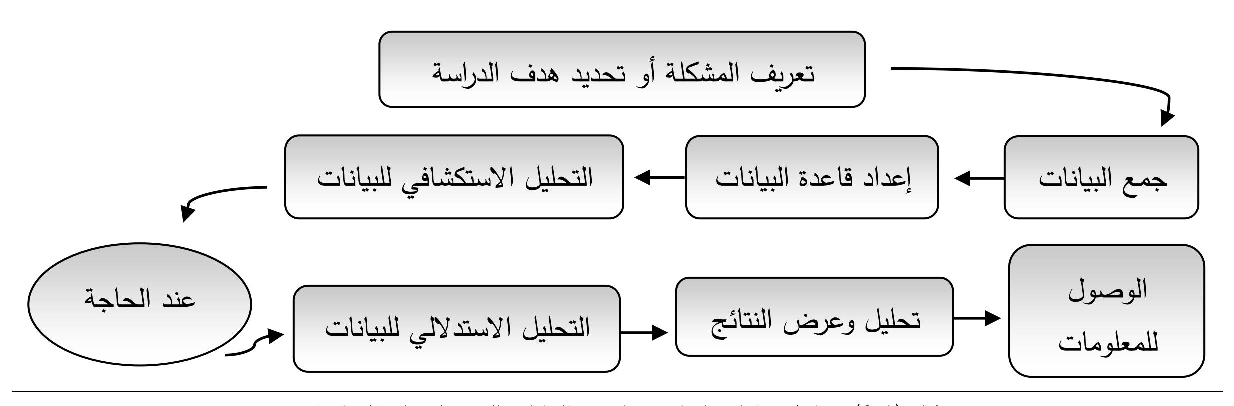
إلى الآن تم في هذا الفصل مناقشة مفهوم البيانات وارتباطه الوثيق بعلم الإحصاء، وكذلك تم توضيح أنواع هذه البيانات ومصادرها وطرق جمعها. نأتي الآن للخطوة التي تلي جمع البيانات، وهي إعدادها وحفظها في قواعد البيانات. إن أول استخدام **لقواعد البيانات** (Databases) بصورة منظمة كان في أواخر الخمسينيات أو أوائل الستينيات (هان

[1] مثال على ذلك إذا ما تخيلنا دراسة عددية تشمل كل سكان العالم والذين هم في حالة تغير دائم بالزيادة (الولادة) أو النقصان (الوفاة) خلال جزء من الثانية.

[2] هنا لا نقصد التطرق لأحد أساليب الإحصاء في الاقتصاد والصناعة وهو ما يعرف **بمراقبة الجودة** (Quality Control)، بل نقصد مفهوم المعاينة بحد ذاته فقط.

(Han)). ومع التطور المتسارع عبر الزمن في الحاسبات والذي صاحبه زيادة غير مسبوقة في أحجام البيانات المرتبطة بالظواهر والدراسات العلمية المختلفة، ظهرت أنواع حديثة من قواعد البيانات والتي أصبحت تعرف **بنظم مستودعات البيانات**[1] (Warehousing Systems). ولكننا هنا لسنا بصدد التحدث عن هذه النظم لأن استخدامها يتطلب الكثير من المهارة والخبرة في التعرف على الأساليب الإحصائية التحليلية المتقدمة والتي لا يتطرق لها هذا الكتاب.

إلا أننا نود القول بأن البيانات التي تتوفر من خلال استخدام طرق جمع البيانات المختلفة وأساليب المعاينة المستخدمة لابد في النهاية من إعدادها في قاعدة بيانات بسيطة كانت (كأن تكون أسماء طلبة ودرجاتهم في أحد الفصول)، أو ضخمة (كتلك البيانات التي تمثل العملية الإنتاجية لسلسلة من المصانع العملاقة خلال عدة سنوات)، وفي كل الأحوال فإن هذا يتم عادة باستخدام الحاسوب[2] سواء كان الغرض هو استخدام الأساليب الإحصائية الاستكشافية أو التحليلية، أو حتى لحفظ البيانات لأغراض إدارية فقط. إن التعامل مع قواعد البيانات بهدف الحصول على معلومات مفيدة لأي مجال من المجالات، التي سبق وأن تكلمنا عنها في هذا الفصل، يتم باستخدام الأساليب الإحصائية من حلال عدة مراحل بحسب الهدف المطلوب تحقيقه من هذه البيانات. والشكل (3.1) يوضح مراحل سير عملية التعلم الإحصائي بصورة مبسطة من خلال خطوات التحليل الاستكشافي والتحليل الاستدلالي الإحصائي.



شكل (3.1): مراحل تعامل علم الإحصاء مع البيانات للحصول على المعلومات.

ونختتم الفصل الأول بمثال، (مثال (5.1))، نوضح من خلاله تطبيق عملي لكيفية تعامل علم الإحصاء، (بدون التطرق لتفاصيل الأساليب الإحصائية المستخدمة)، مع الدراسات المختلفة في الحياة بغية الحصول على المعلومات.

مثال (5.1): أراد قسم الهندسة الميكانيكية بإحدى الجامعات الأمريكية القيام بدراسة إحصائية حول كميات استهلاك الوقود في السيارات المختلفة في السنوات الأخيرة. فتم في البداية، (كما هو موضح في شكل (3.1))، تحديد هدف الدراسة وهو تحديد ودراسة العوامل أو المتغيرات المتعلقة باستهلاك الوقود في السيارات. ثم تم بعد ذلك تحديد مجتمع الدراسة والذي يشمل كل أنواع السيارات المستخدمة في الولايات المتحدة الأمريكية، تم بعد ذلك تحديد المتغيرات التي ستتضمنها الدراسة والتي ستدرج في توصيف البيانات.

---

[1] وهذا ما أدى إلى تطور أحد الجوانب التطبيقية في الإحصاء الحديث بشكل كبير، والذي يعرف حاليا باسم **تنقيب البيانات** (Data Mining).

[2] ومن أشهر برمجيات الحاسوب التي يتم فيها إعداد وتخزين البيانات للمختصين وغير المختصين هي **أوفيس إكسيل** و**أكسس** (Office Excel and Access) وخوادم **إس كيو إل** (SQL Servers). مع إمكانية تخزينها مباشرة في بعض الحزم الإحصائية مثل R، S-plus، SPSS، STATISTICA، وغيرها.

الخطوة التالية كانت تحديد حجم العينة المناسب للدراسة، (وهو $n = 300$ في هذا المثال)، وكانت الاقتراحات تقضي بجمع عينة طبقية نظرا لأن طبيعة كل ولاية أمريكية من حيث المناخ السائد والطبيعة الجغرافية، (طرق منبسطة أو جبلية أو غيرها)، ومدى الازدحام، وغير ذلك من العوامل كان يجب أخذها بالاعتبار. تم بعد ذلك إعداد قاعدة البيانات، والتي يوضحها الجدول (2.1)، بحيث تمثل الأعمدة المتغيرات المستخدمة في الدراسة، و تمثل الصفوف المشاهدات، وعددها هو $n = 300$.

ولاحظ أن المتغيرات الثلاثة الأولى في الجدول، وهي موديل وشكل السيارة ونوع ناقل الحركة هي متغيرات وصفية غير رقمية، وأما باقي المتغيرات وهي عدد الاسطوانات وكمية الوقود المستهلك داخل وخارج المدينة فهي متغيرات متصلة رقمية. وسيكون لكل نوع من المتغيرات الأسلوب المناسب للتعامل معه، كما سنرى في الفصول اللاحقة، سواء خلال مرحلة الاستكشاف أو مرحلة التحليل الاستدلالي إذا ما تطلبت الأهداف التفصيلية للدراسة ذلك.

إن الأساليب الإحصائية التي يمكن تطبيقها، في هذه الدراسة أو غيرها، عادة ما تكون عديدة، ويتطلب التعرف على الطريقة أو الطرق "المثلى" كثيرا من الخبرة التي تكتسب عن طريق كثرة الممارسة للعمل الإحصائي. أما بالنسبة لمرحلة عرض النتائج والحصول على المعلومات، فإنها عادة ما تتطلب تعاونا بين الباحث الأصلي، (وهو قسم الهندسة الميكانيكية)، والباحث الإحصائي الذي قام بجمع وتنظيم وتحليل البيانات. وفي كثير من الحالات، قد يقوم الباحث بإعادة جمع بيانات جديدة إما لتأكيد النتائج التي تحصل عليها، أو بسبب حصوله على معلومات تتعارض مع الطبيعة المنطقية للدراسة.

جدول (2.1): جزء من بيانات خاصة بدراسة أجريت حول استهلاك الوقود في بعض أنواع السيارات.

التوصيف: المتغيرات تمثل مواصفات سيارات من حيث الموديل، الشكل، نوع ناقل الحركة، عدد اسطوانات المحرك، والكمية التقريبية للوقود المستهلك، (بالكيلومتر لكل جالون (20 لتر))، داخل المدينة (سرعات منخفضة)، وخارج المدينة (سرعات عالية).

المشاهدات	المتغيرات					
	موديل السيارة	شكل السيارة	نوع ناقل الحركة	عدد اسطوانات المحرك	استهلاك الوقود داخل المدينة	استهلاك الوقود خارج المدينة
1	Acura NSX	رياضية	أوتوماتيكي	6	27	38
2	Audi A4	صالون صغيرة	عادي	4	35	49
3	Buick Century	صالون متوسطة	أوتوماتيكي	6	32	46
4	Dodge Ram	شاحنة	أوتوماتيكي	8	38	52
⋮	⋮	⋮	⋮	⋮	⋮	⋮
300	BMW 750i	صالون كبيرة	أوتوماتيكي	8	37	48

6.1 تمارين الفصل الأول

تمرين (1.1): إحدى الدراسات الاجتماعية تضمنت 250 أسرة من ذوي الدخل المتوسط في إحدى المدن، وتم تسجيل عدد الأطفال في سن الدراسة تحت 18 سنة لكل أسرة. ما هو المتغير في هذه الحالة وما نوعه؟، وما هو عدد المشاهدات؟.

تمرين (2.1): تم حساب عدد الساعات التي يستغرقها العمل الفعلي لموظفي قسم الحسابات الجارية، والبالغ عددهم 150 موظف، في مجموعة فروع أحد المصارف الدولية. فكانت تتراوح ما بين 3 و 21 ساعة أسبوعيا. وتم كذلك تسجيل الفرع الذي يعمل به كل موظف. ماهي المتغيرات المتوفرة في هذه البيانات وما نوعها؟ وكم هو حجم العينة التي تم اختيارها؟.

تمرين (3.1): نشرت إحدى الصحف تقريرا حول دراسة شملت 320 أسرة في إحدى الدول العربية انتهت قصتها بالطلاق بين الزوجين، وأن 30% من حالات الطلاق في هذه العينة من الأُسر كان السبب الرئيسي فيه هو إدمان أحد الزوجين على متابعة موقع الفيس بوك (Facebook) على شبكة الانترنت. هل ترى أن هذه الدراسة تصنف ضمن الإحصاء الوصفي أم الاستدلالي ولماذا؟.

تمرين (4.1): أي من المتغيرات التالية هو متغير وصفي (قطاعي)؟

أ. الزمن المستغرق لبناء قاعدة اسمنتية في 10 مواقع بناء.
ب. تصنيف المدارس إلى حكومية وخاصة في مدن جمهورية مصر العربية.
ج. رأي مجموعة من الأشخاص في أداء الحكومة السابقة في بلدانهم ما إذا كان؛ سيء، مقبول، جيد، أو ممتاز.
د. نتيجة إجراء الكشف الطبي للمرضى في أحد المستشفيات التي تضمنت؛ سليم، مصاب بالفيروس، حامل للفيروس.

تمرين (5.1): وضح نوع المقياس المستخدم في رصد البيانات التالية:

أ. الأصل العرقي لسكان شمال أفريقيا. ب. التكلفة الشهرية لخدمة الانترنت بحد أقصى 40 جيجا بايت في دولة ما. ج. الرتب العسكرية المتعارف عليها في الجيش. د. التقويم السنوي معطى بالأرقام.

تمرين (6.1): قم بإعداد جدول بيانات لدراسة افتراضية (اقتصادية، اجتماعية، طبية، ...) مستخدما حجم عينة صغير نسبيا. يمكنك الاستعانة بجدول (2.1).

الفصل الثاني

التحليل الاستكشافي للبيانات – الجزء الأول

(Exploratory Data Analysis (EDA) – Part One)

1.2 مفهوم الإحصاء الاستكشافي (Concept of Exploratory Statistics)

2.2 توزيع البيانات وجداول التوزيع التكراري (Data Distribution and Frequency Tables)

1.2.2 توزيع البيانات (Data Distribution)

2.2.2 الجداول التكرارية (Frequency Tables)

3.2 التمثيل البياني للبيانات (Graphical Display of Data)

1.3.2 الرسم النقطي (Dot or Point Diagram)

2.3.2 مخطط الزمن (Time Chart)

3.3.2 المدرج التكراري (Histogram)

4.3.2 المضلع التكراري (Frequency Polygon)

5.3.2 المضلع التكراري المتجمع (Frequency Ogive)

6.3.2 الأعمدة البيانية (Bar Charts)

7.3.2 القطاعات الدائرية (Pie Charts)

1.2 مفهوم الإحصاء الاستكشافي (Concept of Exploratory Statistics)

ناقشنا في الفصل الأول كيف أن كل مجموعة من البيانات قد تحتوي على معلومات هامة كامنة بداخلها، وأن استخدام الإحصاء كخطوة أولى، بعد مرحلة جمع البيانات، يكون عادة بإعطاء ملخص عن هذه البيانات من خلال بعض المقاييس الرقمية والرسومات البيانية التوضيحية للحصول على معلومة سريعة أو فهم لما تمثله البيانات، ومدى دقتها ومنطقيتها. وقد يكون هذا الفهم كافيا بحد ذاته للوصول للهدف المطلوب، أو قد يتعدى الأمر إلى المزيد من التحليل الاستدلالي العميق، والذي بناءا على تعريف المجتمع والعينة، يمكن النظر إليه على أنه ذلك الفرع الذي يحوي الأساليب التي تستخدم لتكوين استنتاجات حول المجتمع اعتمادا على المعلومات المتوفرة في العينة.

من ناحية أخرى، فإن استخدام الإحصاء بصورة "صحيحة" يؤدي بدوره للوصول إلى الأهداف أو المعلومات المرجوة، أما في حالة الاستخدام "الخاطئ" أو الغير دقيق فقد نحصل على جزء فقط من المعلومات أو قد نلاحظ في بعض الدراسات الإحصائية، أو الدراسات الأخرى التي تحوي جانبا إحصائيا، أن النتائج لا تأتي بمعلومات ذات قيمة أو لا تصلح لاتخاذ القرار المناسب.

ومن الأسباب التي تؤدي لذلك هو إهمال المرحلة التي تهتم باستكشاف البيانات، أو على الأقل عدم إعطاء هذه المرحلة الاهتمام الكافي نتيجة قلة الخبرة في التعامل مع البيانات. ولأنك "تحصد" (نتيجة التحليل) ما "تزرع" (البيانات)، أو كما يقول المثل الأمريكي الشهير؛ "(GIGO: Garbage In, Garbage Out)"، والذي يعني أنه إذا كانت المدخلات (البيانات) هي مجرد "نفايات"، فالمخرجات (النتائج) ستكون نفايات أيضا، فإن مرحلة التحليل الاستكشافي للبيانات يجب أن تأخذ نصيبا وافرا من الوقت والتركيز لتسليط الضوء على مدى **مصداقية البيانات** (Data Reliability) قبل الانتقال للخطوة التالية. فمثلا، لا يمكن الاعتماد على عينة مكونة من مشاهدات تم جمعها من درجات طلبة "فصل المتفوقين" في أحد المدارس الإعدادية النموذجية لوصف مستوى طلبة المرحلة الإعدادية في تلك الدولة.

إن الهدف من استخدام التحليل الاستكشافي للبيانات، في الواقع العملي، هو تمكين الباحث من استيعاب وفهم "ما تقوله" البيانات وما يمكن أن تحتويه من معلومات، وخاصة عندما تكون قاعدة البيانات ضخمة ومعقدة. وهذا الهدف يمكن تحقيقه من خلال تلخيص مفردات البيانات بشكل يبرز **النمط** (Pattern) أو الأنماط السائدة في هذه البيانات بعيدا عن التغيرات العشوائية التي قد تحدث، ويقصد بالنمط هنا التغير المنتظم في البيانات. فمثلا، إذا ما كانت درجات أحد الطلبة الجامعيين خلال 7 فصول دراسية في بعض المقررات هي على الترتيب: 61، 63، 67، 70، 71، 72، 50 فنلاحظ وجود تغير منتظم في درجات هذا الطالب يمثله الارتفاع المتدرج في الدرجات خلال الستة فصول الأولى، وهذا هو النمط السائد في هذه البيانات، ثم نلاحظ انخفاض حاد في درجة مقرر في الفصل السابع وهو يعد تغيرا عشوائيا يسمى أحيانا **بالإزعاج** أو **الضجيج** (Noise).

وللقارئ أن يتخيل صعوبة تتبع كلا من الأنماط و/أو الإزعاج في قاعدة بيانات تحتوي آلاف أو ملايين المفردات، إذ أن المثال التوضيحي السابق يمثل متغيرا واحدا يضم 7 مفردات فقط، أما عمليا فإننا نتعامل عادة مع أكثر من متغير، وتلك المتغيرات تضم عددا كبيرا جدا من المشاهدات، ومما يزيد الأمر تعقيدا نشوء علاقات بين المتغيرات بعضها مع البعض، وبين المشاهدات مع بعضها، وكذلك بين المتغيرات والمشاهدات، وهذا كله

يؤكد على ضرورة استخدام التحليل الاستكشافي للبيانات للوقوف على طبيعة تلك العلاقات وسبر أغوارها لفهم سلوك البيانات. إلا أنه يجب التأكيد على أنه لا يمكن بناء قرار حاسم بدون تدخل التحليل الاستدلالي، إلا في تلك الحالة التي يتم فيها استخدام كل مفردات مجتمع الدراسة، وهي حالة نادرة عمليا. إلا أن هذا لا يقلل من أهمية الإحصاء الاستكشافي أو الوصفي، حيث أنه إضافة لما سبق فإن استخدامه قد يفتح لنا آفاق أو أفكار جديدة لاستخلاص معلومات إضافية من البيانات، معلومات لم تكن مدرجة في بنود الدراسة الأصلية.

إن الغرض من استخدام التحليل الإحصائي الاستكشافي لوصف البيانات لا يعني بالضرورة أن نكون مهتمين بوصف كل مفردة من مفردات البيانات، لأن الخلاصة التي قد نخرج بها لا تعتمد عموما على قيم البيانات "الخام" بحد ذاتها، بل تعتمد على ما تحتويه هذه البيانات ككل. فمثلا، عند القيام بتجربة لدراسة مدى صلاحية المياه للشرب في مدينة بنغازي وذلك عن طريق اختيار 50 عينة من المياه من مناطق مختلفة من المدينة، فنحن في هذه الحالة لسنا مهتمين بالقراءات الـ 50 بحد ذاتها بل بما يمكن أن تمنحه لنا من خلاصة أو معلومات للإجابة على التساؤل المطروح.

إن أدوات التحليل الاستكشافي، كما سنرى في هذا الفصل، هي عبارة عن مجموعة من الإحصاءات التلخيصية (الرقمية) والرسومات التوضيحية البيانية والتي يتم استخدامها لتنظيم وتلخيص ومن ثمة تحويل قاعدة البيانات بما تحويه من مشاهدات متناثرة (عددية أو وصفية) إلى صورة رقمية أو مرئية تكون مفهومة للمتخصصين وغير المتخصصين.

وهكذا، بعد هذه المناقشة يمكننا تعريف الإحصاء الاستكشافي (أو الوصفي)، والذي يتم تنفيذه "حديثا" من خلال مرحلة التحليل الاستكشافي للبيانات بالصورة التالية:

تعريف (1.2): **تحليل البيانات الاستكشافي** (Exploratory Data Analysis, (EDA)): هو ذلك الفرع في الإحصاء والذي يشتمل على الطرق الرقمية والبيانية التي تهتم بجمع وتنظيم واستكشاف وعرض الصفات المميزة ضمن البيانات بغية الحصول على المعلومات الكامنة بداخلها.

وفي نهاية هذا الجزء سنقوم بعرض أهم الخواص التي يفضل توفرها في المقاييس الإحصائية الاستكشافية:

- أن يكون المقياس عبارة عن قيمة واحدة معبرة، حيث أن الهدف هو تلخيص ووصف البيانات باختصار.
- سواء كانت طبيعة البيانات عددية أو وصفية، يفضل أن يكون المقياس قابلا للاستخدام في العمليات الجبرية الرياضية المختلفة وكذلك عمليات التحويل.
- يفضل في حساب المقياس أن يتضمن كل مفردات البيانات، لأن عملية وصف البيانات تعتمد على التلخيص، وهذا التلخيص قد يؤدي لفقد لبعض المعلومات. هذا الفقد قد يتقلص إذا ما تم استخدام كل المفردات في حساب المقياس.

وفي البيانات التي تحتوي على تكرار للمفردات، يفضل أن يأخذ المقياس هذا التكرار بعين الاعتبار، وهذا ما سنلاحظه لاحقا خلال هذا الفصل.

2.2 توزيع البيانات وجداول التوزيع التكراري (Data Distribution and Frequency Tables)

1.2.2 توزيع البيانات (Data Distribution)

بعد تجميع البيانات من مصادرها، تبرز الحاجة لوضعها في قاعدة بيانات عامة، (في حالة عدم وجود هدف مسبق أو دراسة محددة سلفا)، أو خاصة من أجل تحقيق الهدف المطلوب. وفي كثير من الحالات، نكون بحاجة لتلخيص مفردات البيانات بغرض عرضها (وصفها) أو تحليلها. ونعني بذلك تحويلها من مفردات "خام" إلى مفردات منظمة في جداول. ولتوضيح أهمية تحويل البيانات من مجرد قيم متناثرة إلى نظام مكون من صفوف وأعمدة دعنا نتناول المثال التوضيحي الشهير التالي (فرانك (Frank)): إذا حاولت أن تتخيل نقطة واحدة في الفراغ داخل عقلك، ثم تخيلت نقطة ثانية وثالثة ... وهكذا، فإنك ستفقد تركيزك في أماكن هذه النقاط سريعا ربما بعد النقطة الخامسة بشرط عدم تخيل هذه النقاط ضمن صورة هندسية منظمة، (خط مستقيم، مربع،...). وهذا يشابه الفكرة الأساسية في وضع مفردات البيانات داخل نظام معين، والذي ينتج عنه ما يعرف **بتوزيع البيانات** (Data Distribution). فتوزيع البيانات هو تلخيص لقيم المفردات لمتغير يوضح نطاق هذه القيم (غالبا من الأقل إلى الأكبر)، وتكراراتها. وهنالك طريقتان لوصف تكرار أو تعدد القيم ضمن توزيع البيانات:

الطريقة الأولى هي طريقة **التكرار** (Frequency)، وهي تمثل عدد مرات وجود كل قيمة (مفردة) من قيم البيانات، وهذه الطريقة هي الأكثر استخداما في تلخيص البيانات وتعتمد على أن عدد المشاهدات معلوم. أما الطريقة الثانية فهي طريقة **التكرار النسبي** (Relative Frequency)، وتعتمد على حساب نسبة وجود كل مفردة وذلك بقسمة عدد مرات تكرارها على عدد المفردات الكلي. وهنالك في الواقع بعض الخصائص المميزة لتوزيع البيانات والتي يجب على الباحث أو مستخدم التحليل الاستكشافي التحقق منها، بغض النظر عن أهداف الدراسة، أهمها هي:

- **مركز البيانات** (Data Centre): وهي القيمة التي تقع في منتصف قيم البيانات، حيث تتركز القيم الأكثر انتشارا، (كما سنرى في الجزء الخاص بمقاييس النزعة المركزية (4.2)).
- **انتشار البيانات** (Data Spread): والذي يصف مقدار التغير في قيم المفردات حول مركز البيانات. فكلما اقتربت القيم من مركز البيانات كان انتشارها أقل، والعكس بالعكس، (كما سنرى في الفصل الثالث الخاص بمقاييس التشتت (1.3)).
- **شكل البيانات** (Data Shape): والذي يساعدنا في الوقوف على سمات عدة لتوزيع البيانات، ويتم الوصول إليه من خلال الجداول التكرارية والرسومات البيانية، (وسنقوم بدراسة الجداول التكرارية في الجزء الحالي، أما الرسومات البيانية فسنتناولها في الجزء التالي (3.2)).

وقبل أن نتناول موضوع الجداول التكرارية بالشرح المفصل، نود تسليط الضوء على أهمية دراسة شكل البيانات ودورها في فهم توزيع البيانات من خلال النقاط التالية:

- دراسة شكل البيانات يساعدنا في تحديد مواضع تمركز البيانات، ومراقبة ما إذا كان لها أكثر من مركز، وملاحظة مدى انتشارها بصورة متماثلة[1].

---

[1] سيتم توضيح مفهوم التماثل في توزيع البيانات لاحقا في هذا الفصل.

- يمكن من خلال توزيع البيانات تحديد أماكن **الفجوات** (Gaps)، وهي ترمز لعدم وجود قيم معينة[1] ضمن نطاق المفردات. فمثلا عدم وجود التقدير "جيد جدا" ضمن مفردات بيانات تمثل تقديرات مجموعة كبيرة من الطلبة يعني وجود فجوة في البيانات، وقد يكون مؤشرا على وجود خلل في عملية تقييم الطلبة.
- شكل البيانات يمكن أن يساعد أيضا في استكشاف القيم **المتطرفة** أو **الشاذة** (Outliers) وهي المفردات التي تشذ عن النمط السائد في قاعدة البيانات، سواء بكونها تأخذ قيما صغيرة جدا أو كبيرة جدا. فمثلا، وجود طلبة نظاميين مسجلين في كلية الآداب أعمارهم تفوق الـ 30 عاما يعبر عن وجود قيم متطرفة في قاعدة البيانات التي تضم أعمار الطلبة، مع متغيرات أخرى، في كلية الآداب. وظهور القيم المتطرفة في البيانات يكون عادة لعدة أسباب منها أخطاء في رصد المفردات، أو أن هذه المفردات قد تم سحبها من خارج مجتمع الدراسة، أو ببساطة وجود خلل في الظاهرة محل الدراسة، وغير ذلك من الأسباب. ولابد من الإشارة هنا إلى أن وجود القيم المتطرفة لا يعد "عيبا" في جميع قواعد البيانات، إذ أنه في بعض الحالات قد تكون التجربة أو الدراسة مرتكزة على إظهار تلك القيم وتحليل أسباب وجودها وإيجاد الحلول. فمثلا، في بعض الأبحاث الطبية قد يكون من المفيد الوقوف على تلك القيم المتطرفة والتي تمثل فئات عمرية صغيرة (أطفال) عند دراسة خاصة بمرض السكر.

2.2.2 الجداول التكرارية (Frequency Tables)

نأتي الآن لعرض طريقة التعامل مع **الجداول التكرارية**، والتي تستخدم لتلخيص وتنظيم البيانات بدون أن نفقد أي معلومات موجودة ضمن البيانات لأن هذه الجداول لا تختزل المفردات بل تنظمها فقط. وهذه الجداول، والتي تسمى أيضا **بجداول التوزيع التكراري** (Frequency Distribution Tables)، تأخذ عدة أشكال بحسب نوع البيانات أو طبيعة قيم المفردات. فيمكن استخدامها مع البيانات العددية أو الوصفية على حد السواء، وكذلك يمكن عرض المفردات بقيمها الفعلية أو من خلال تكوين فترات كما سنرى من خلال عرض الأمثلة التالية:

مثال (1.2): إذا كانت النتائج الفصلية لـ 15 طالبا من طلبة قسم الرياضيات للفصل الماضي هي A, B, B, B, A, C, F, F, D, A, D, D, B, D, C فإن أول ما نلاحظه أن البيانات تمثل قيم وصفية متناثرة لا تعبر عن مستوى الطلبة بشكل مباشر أو صريح، وكلما كان عدد الطلبة (المشاهدات) أكبر كلما كان الوضع أكثر غموضا. لذلك، فإننا نقوم بتلخيص هذه المشاهدات في جدول، (جدول (1.2))، والذي يمثل فيه العامود الأول المتغير X ، (تقدير الطالب)، مرتبا من التقدير الأعلى إلى الأقل، مع إمكانية اعتبار عكس ذلك، والعامود

جدول (1.2): الجدول التكراري لتقديرات الطلبة في المثال (1.2)

المتغير X (تقديرات الطلبة)	التكرار f
A	3
B	4
C	2
D	4
F	2
المجموع	15

[1] لابد من التفريق هنا بين مفهوم "الفجوة" ومفهوم "**القيم المفقودة**" في البيانات، حيث أن الثانية تعني وجود نقص ضمن عدد المفردات الكلي.

الثاني يمثل تكرار حدوث المفردة، والذي يرمز له عادة بالرمز f. فيتم حساب عدد مرات تكرار كل تقدير وكتابته في العامود الثاني، مقابل التقدير المناظر في العامود الأول. ولاحظ أن مجموع التكرارات لابد أن يساوي، دائما، مجموع مفردات البيانات وهو في هذا المثال يساوي 15. ويسمى الجدول (1.2) بجدول التوزيع التكراري لتقديرات طلبة الرياضيات.

مثال (2.2): قام أحد الباحثين بإجراء دراسة اجتماعية على 25 مريضا بحيث تم تسجيل عدد الزوار، (ويمثله المتغير X)، لهم خلال أول ثلاثة أيام من دخولهم المستشفى فكانت النتيجة 2، 2، 1، 2، 2، 1، 3، 3، 2، 3، 1، 4، 3، 3، 2، 2، 1، 3، 2، 1، 1، 1، 0، 2، 2. يمكننا الآن إعداد جدول توزيع تكراري لهذه البيانات (جدول (2.2)) بنفس الكيفية السابقة عن طريق وضع المتغير (عدد الزوار) في العامود الأول، والذي سيتم عرضه في الجدول التكراري في الصف الأول وتكرارات الحدوث في العامود الثاني، والذي سيعرض في الصف الثاني في الجدول المرفق، لكي يعتاد القارئ على التعامل مع كلتا طريقتي العرض، العمودية والأفقية.

جدول (2.2): الجدول التكراري لتوزيع عدد الزوار في المثال (2.2)

المتغير X (عدد الزوار)	0	1	2	3	4	المجموع
التكرار f	1	7	10	6	1	25
التكرار النسبي $\frac{f}{N}\times 100$	4%	28%	40%	24%	4%	100%

ولاحظ من الجدول أن المتغير X لا يأخذ قيما متكررة بل يتم حساب التكرار لكل قيمة مناظرة له، علما بأن القيمة 0 تعني عدم حضور زوار للمريض. أما الصف الثالث في الجدول فهو يمثل التكرار النسبي والذي يتم حسابه عن طريق قسمة قيمة كل مفردة من مفردات المتغير X على مجموع المفردات $N = 25$. ولاحظ أيضا أن المتغير X يأخذ قيما عددية صحيحة تتراوح بين 0 و 4، وهذا ما جعل الجدول (2.2) يظهر بصورة مقبولة عمليا، أما في الحالات التي تكون فيها قيم المفردات أكثر "تنوعا"، (إذا كان عدد الزوار أكثر من 4 زوار بكثير)، فإن شكل الجدول التكراري لن يكون عمليا ولن يؤدي الغرض الأساسي منه وهو تلخيص البيانات. لهذا نقوم بتجميع المفردات بصورة **فئات** أو **فترات** (Intervals or Classes)، وحساب تكراراتها كما سنرى في المثال التالي:

مثال (3.2): لنفرض أن الدراسة في المثال السابق (2.2) انبثقت عنها دراسة أخرى تهتم بتكلفة العلاج في المستشفيات الخاصة في ليبيا في السنوات الأخيرة، فتم تسجيل تكلفة العلاج بالدينار الليبي (المتغير X) لعدد 1000 مريض (عدد المشاهدات). عندئذ سيكون لدينا بيانات تتألف من ألف قيمة، (والتي لا داعي لسردها بالتفصيل في هذا المثال)، يمكن تلخيصها باستخدام الفترات كما هو موضح في الجدول (3.2). ويمكننا، على سبيل المثال، ملاحظة أن قيمة العلاج "لأكثر" المرضى يتراوح ما بين (30 – 69) دينار ليبي.

جدول (3.2): التوزيع التكراري لتكلفة العلاج لعدد 1000 مريض في المستشفيات الخاصة في ليبيا.

المتغير X (تكلفة العلاج)	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
التكرار f	13	41	93	147	240	200	160	67	13	26

نأتي الآن لوضح الخطوات الرئيسية في تكوين جدول التوزيع التكراري من البيانات المفردة لأي متغير:

1) نقوم بتحديد أصغر قيمة، (ولتكن $x_{\min}$)، وأكبر قيمة، (ولتكن $x_{\max}$)، ضمن مشاهدات المتغير.
2) نقوم بتحديد عدد الفئات أو الفترات المناسب، (والذي سنرمز له بالرمز c). وهذا التحديد يتم عادة بصورة تقديرية بحيث يتناسب طرديا مع حجم المشاهدات وطبيعة قيم تلك المشاهدات، وهو أمر يرجع تقديره للباحث أو الدارس. إلا أنه يجب التنويه إلى وجود قاعدة رياضية تعرف بقاعدة **ستيرجس** (Sturges Rule)، والتي يمكن لحديثي التعامل مع الجداول التكرارية استخدامها. هذه القاعدة لها الصورة التالية:
$c = 1+\log_{10}(2)\times\log_{10}(N) = 1+(3.32)\times\log_{10}(N)$ ، حيث N هو عدد المشاهدات، ويتم تقريب قيمة c إلى أقرب عدد صحيح.
3) نقوم بحساب الفرق بين أصغر قيمة وأكبر قيمة في المشاهدات، وهو ما يعرف **بالمدى**[1] (Range)، والذي سنرمز له بالرمز R، فيكون $R = x_{\max} - x_{\min}$.
4) يتم حساب طول الفترة (L) بالصورة $L = R/c$.

هذه الخطوات هي في الواقع الخطوات العامة لتكوين جدول توزيع تكراري يعرف **بالمنتظم** لأن طول الفترات فيه يكون متساويا، وهذا النوع من الجداول هو الأكثر استخداما من الناحية العملية، أما في بعض الحالات التي تكون فيها المشاهدات "مبعثرة" بشكل كبير بحيث تكون بعض الفترات خالية من التكرارات فإنه من الأفضل عندئذ تنسيق طول الفترات بحيث يتناسب مع هذه الوضعية، ويعرف التوزيع التكراري في هذه الحالة بأنه **غير منتظم**.

وتوجد أيضا حالات أخرى يصعب فيها تحديد الحد الأدنى و/أو الحد الأعلى لقيم المشاهدات، وذلك عند عدم معرفة الباحث بتلك الحدود، (كما هو الحال في الأبحاث المرتبطة بالتحديث المستمر لقاعدة البيانات على الإنترنت مثل بيانات الأرصدة المصرفية، أو الدراسات مجهولة القيم الدنيا والعليا كما هو الحال في بعض التجارب الكيميائية)، فيتم عندئذ إعداد الجدول التكراري بحيث يكون مفتوح الطرفين (الحد الأدنى والحد الأعلى)، أو مفتوح عند أحدهما فقط.

من ناحية أخرى، فإن طبيعة المتغير محل الدراسة قد تفرض علينا نوعان من جداول التوزيع التكراري؛ الجداول التكرارية **المستمرة** أو **المتصلة** (Continuous Frequency Tables)، والجداول التكرارية **الغير مستمرة** (Discontinuous Frequency Tables). فالمتغيرات عندما تأخذ مفرداتها قيما مستمرة، مثل الأجور والأوزان والأحجام وغيرها، يكون من الأفضل استخدام الجداول التكرارية المستمرة معها، فمن الناحية العملية، يتم حساب أجر الموظف، بالدينار الليبي مثلا، بعدد صحيح وكسر عددي ، فإذا كانت الفترة في الجدول هي (300 – 500) والتي تليها (501 – 701)، فيكون الموظف الذي أجره 500.520 دينارا خارج الفترتين.

ولتوضيح كيفية تكوين جدول التوزيع التكراري عمليا لنأخذ المثال التالي:

[1] سيتم التطرق لمفهوم المدى كمقياس للتشتت بالتفصيل لاحقا في الفصل القادم.

مثال (4.2): البيانات التالية (جدول (4.2)) تمثل أوزان 40 طفلا بالكيلوجرام بعد مرور شهر على الولادة، والذين تم تسجيلهم في برنامج المتابعة الدورية، في أحد المستشفيات الحكومية.

جدول (4.2): أوزان 40 طفلا (بالكجم) في أحد المستشفيات الحكومية.

3.6	4.0	4.7	4.2	5.5	4.6	5.1	3.2
4.7	5.7	4.1	4.8	4.3	4.2	2.6	4.4
4.1	4.9	4.3	3.9	4.6	4.4	5.3	3.5
4.4	2.9	5.1	4.2	5.4	4.7	4.1	4.3
4.5	5.2	4.0	4.9	3.6	4.2	4.8	5.7

وكان المطلوب هو إنشاء جدول توزيع تكراري، فنقوم بإتباع خطوات تكوين جدول التوزيع التكراري السابقة، فيكون لدينا:

1) أصغر قيمة من قيم المشاهدات في الجدول هي x_{min}= 2.6، وأكبر قيمة هي x_{max}= 5.7 .
2) بالنظر لعدد المشاهدات $N = 40$، نرى أن اختيار عدد فترات $c = 7$ سيكون مناسبا. (يمكننا أيضا استخدام قاعدة استيرجس هنا فنحصل على $c = 1+(3.32)\times\log_{10}(40) = 6.29$ ، والذي بعد التقريب، لأعلى قيمة، يكون مساويا لعدد الفترات الذي تم اختياره بدون استخدام القاعدة).
3) نقوم بحساب المدى $R = x_{max} - x_{min} = 5.7 - 2.6 = 3.1$.
4) نحسب طول الفترة، $L = R / c = 3.1/7 = 0.44$ ، ونعتبر التقريب $L = 0.5$ ، وذلك لإعطاء المشاهدات مجالا أوسع لكي تقع قيمها ضمن نطاق الفترات.

يمكننا الآن البدء بتكوين أول فترة في الجدول التكراري والتي يمكن أن يكون الحد الأدنى لها هي أصغر قيمة في المشاهدات 2.6، كما يمكن البدء بقيمة أقل منها قليلا مثل القيمة 2.5، وعلى هذا فإن أول فترة في الجدول ستكون (2.5 – 2.9)، حيث أن طول الفترة[1] هو 0.5، وعلى فرض أننا سنقوم بإنشاء جدول توزيع تكراري منتظم وغير مستمر، ستبدأ الفترة الثانية بالقيمة 3.0، وبإضافة طول الفترة 0.5 إلى تلك القيمة، مع ملاحظة عد 3.0 كأول قيمة، تكون الفترة الثانية (3.0 – 3.4)، ...، وهكذا إلى أن نحصل على آخر فترة، والتي تشمل على أكبر قيمة في المشاهدات، وهي (5.5 – 5.9). ونلاحظ أن العامود الثاني من اليمين في الجدول (5.2) يحتوي على هذه الفترات مرتبة من الأقل إلى الأعلى طوليا.

جدول (5.2): جدول التوزيع التكراري لأوزان الأطفال المسجلين.

الفترة	التكرار f
2.5 - 2.9	2
3.0 - 3.4	1
3.5 - 3.9	4
4.0 - 4.4	15
4.5 - 4.9	10
5.0 - 5.4	5
5.5 - 5.9	3
المجموع	40

بعد ذلك نقوم بحساب تكرار المشاهدات المناظر لكل فترة، ويتم ذلك بترتيب المشاهدات تصاعديا ثم تحديد عدد المشاهدات التي وقعت في أول فترة، فنرى أن المشاهدتين 2.6 و2.9 تقعان ضمن نطاق الفترة الأولى، وبالتالي فإن تكرار الفترة الأولى يكون $f = 2$ ، وهكذا حتى الانتهاء من حصر كل المشاهدات وتسجيل تكراراتها في الفترات المناظرة كما هو موضح في جدول (5.2).

ولاحظ أن البيانات الأصلية كانت مسجلة باعتبار خانة عشرية واحدة بعد الفاصلة، وبالتالي كانت المسافة بين القيمة العليا لأي فترة والقيمة

[1] حيث أن وحدة العد (التغير) هي 0.1 كيلوجرام، فنعتبر أن 2.5 هي أول قيمة في الفترة ثم يتم إضافة 0.1 لها فنحصل على القيمة الثانية 2.6 وهكذا حتى القيمة الخامسة وهي 2.9 .

الدنيا للفترة التي تليها هي 0.1 . أما إذا كانت المشاهدات مسجلة بأكثر من خانة عشرية واحدة، عندها قد تقع بعض القيم في المسافة "المنفصلة" بين الفترات. لذلك فإنه من الأفضل عادة تكوين فترات مستمرة، كما أشرنا سابقا، بالصورة التالية:

نقوم بقسمة المسافة 0.1 على 2، فنحصل على 0.05 ، ثم يتم طرح هذه القيمة من الحدود الدنيا للفترات وإضافتها إلى الحدود العليا للفترات فنحصل على فترات مستمرة كما هو موضح في الجدول (6.2).

جدول (6.2): جدول التوزيع التكراري لأوزان الأطفال متضمنا مراكز الفترات والتكرار النسبي.

التكرار النسبي	مركز الفترة x	التكرار f	الفترات (المستمرة)	الفترات (الغير مستمرة)
0.05	2.7	2	2.45 - 2.95	2.5 - 2.9
0.025	3.2	1	2.95 - 3.45	3.0 - 3.4
0.1	3.7	4	3.45 - 3.95	3.5 - 3.9
0.375	4.2	15	3.95 - 4.45	4.0 - 4.4
0.25	4.7	10	4.45 - 4.95	4.5 - 4.9
0.125	5.2	5	4.95 - 5.45	5.0 - 5.4
0.075	5.7	3	5.45 - 5.95	5.5 - 5.9
1		40	المجموع	

وهنا يجب التنويه إلى أنه في جداول التوزيع التكراري التي تحتوي فترات لا يكون هنالك وجود فعلي للمتغير الذي يمثل المشاهدات، (وهي أوزان الأطفال في المثال الحالي)، كما هو الحال في الأمثال السابقة، لذلك يتم في هذه الحالة تمثيل هذا المتغير عن طريق ما يعرف **بمركز الفترة** (Class Midpoint)، والذي يتم حساب قيمته لكل فترة عن طريق جمع الحد الأدنى والحد الأعلى للفترة (المستمرة أو غير المستمرة) وقسمة المجموع على 2 . وهكذا نحصل على عامود جديد في الجدول التكراري يمثل التغير في المشاهدات. في المثال الحالي هذا العامود هو الثالث من اليسار في جدول (6.2)، أما العامود الرابع من اليسار فهو يمثل التكرار النسبي والذي تم حسابه، كما شرحنا سابقا، بقسمة[1] قيمة كل تكرار f على عدد المفردات الكلي $N = 40$.

رأينا في ما سبق كيف أن الهدف في حساب التكرار هو تلخيص البيانات على شكل جدول أكثر تنظيما مما يجعلها أكثر سهولة وبساطة في عملية الاستكشاف الإحصائي والوصف والتحليل وأيضا العرض. وهنالك بعض المقاييس الاستكشافية أو الوصفية التي سنتعرض لها في هذا الفصل والتي تعتمد على إنشاء ما يعرف **بالتكرار المتجمع الصاعد** (Upper Cumulative Frequency, U.C.F)، و**التكرار المتجمع الهابط** (Lower Cumulative Frequency, L.C.F). ويتم حساب المقياس الأول عن طريق تجميع التكرارات الأقل من الحد الأعلى المناظر لكل فترة على حده، أما التكرار المتجمع الهابط فيتم حسابه بالعكس أي بطرح التكرارات الأكثر من الحد الأدنى المناظر لكل فترة على حده. والجدول (7.2) يوضح آلية الحساب للتكرار المتجمع الصاعد

[1] هذا العامود يمكن كتابته بصيغة النسبة المئوية بالطبع إذا ما تم ضرب قيمه في 100 .

والهابط باستخدام بيانات المثال (4.2). ويمكن أيضا استخدام التكرار النسبي لحساب التكرار المتجمع إذا ما دعت الحاجة لذلك.

جدول (7.2): جدول التوزيع التكراري لأوزان الأطفال في المثال (4.2) متضمنا التكرار المتجمع.

التكرار المتجمع الهابط		التكرار المتجمع الصاعد			
التكرار	الفترة	التكرار	الفترة	التكرار f	الفترات (المستمرة)
40	أكثر من 2.45	0	أقل من 2.45		
38	أكثر من 2.95	2	أقل من 2.95	2	2.45 - 2.95
37	أكثر من 3.45	3	أقل من 3.45	1	2.95 - 3.45
33	أكثر من 3.95	7	أقل من 3.95	4	3.45 - 3.95
18	أكثر من 4.45	22	أقل من 4.45	15	3.95 - 4.45
8	أكثر من 4.95	32	أقل من 4.95	10	4.45 - 4.95
3	أكثر من 5.45	37	أقل من 5.45	5	4.95 - 5.45
0	أكثر من 5.95	40	أقل من 5.95	3	5.45 - 5.95
				40	المجموع

3.2 التمثيل البياني للبيانات (Graphical Display of Data)

ناقشنا في الجزء السابق كيفية تلخيص وتبويب البيانات في صورة منظمة تمهيدا لاكتشاف ووصف ما يمكن أن تحويه هذه البيانات من معلومات، إلا أن الجداول التكرارية قد لا تكفي أحيانا بمفردها للتوضيح والوصف وخاصة أن بعض غير المتخصصين في الإحصاء أو المبتدئين قد يواجهون صعوبة في إدراك ما تصفه هذه الجداول والأرقام "التقليدية". لهذا فإننا نكون دائما، أثناء العمل الإحصائي، بحاجة لاستخدام أسلوب، أو إذا صح التعبير، استخدام "لغة" الرسم أو التمثيل البياني الإحصائي لعرض البيانات والنتائج ووصفها بصورة ملخصة. حيث أن هذه الرسوم هي أسلوب عملي يساعد في تكوين فكرة سريعة ودقيقة عما تحويه قواعد البيانات، التي كثيرا ما تكون معقدة، من معلومات كامنة.

إضافة إلى ذلك، فإن بعض هذه الرسوم البيانية تشكل في الواقع الخطوة الأولية أو الاختبارية التي تسبق تطبيق بعض التحليلات الإحصائية المتقدمة والتي ترتكز على فرضيات يتم اختبارها عن طريق الرسم البياني[1]. وقد تختلف طرق العرض البياني، (أو كما يروق لكثير من الإحصائيين تسميته بالعرض البصري (Visualization) في المنشورات الحديثة)، باختلاف طبيعة البيانات أو أهداف البحث، حيث أن التمثيل البياني يجب تصميمه بحيث يبرز الفكرة الرئيسية التي يرغب الباحث بعرضها وتحليلها. ولهذا، تكون الخطوة الأولى عند استخدام التمثيل البياني هي اختيار نوع الرسم (الأسلوب) الذي يناسب طبيعة ونوع البيانات وفي نفس الوقت يبرز الفكرة أو يصف الظاهرة المراد دراستها أو وصفها.

[1] من أشهر الأمثلة على ذلك مفهوم **تحليل الانحدار** (Regression Analysis) في الاستدلال الإحصائي.

إن المقولة المشهورة "صورة واحدة تساوي ألف كلمة"، يتماشى إلى حد كبير مع المنطق الإحصائي في تمثيل البيانات، حيث أن البيانات يتم فهمها ووصفها بشكل أوضح وأسرع إذا ما تم عرضها على هيئة رسوم بيانية، بل يمكننا القول أن صورة واحدة قد تساوي ألف مفردة أو أكثر من مفردات البيانات طالما أن هذه الصورة قد تم "التقاطها" بشكل صحيح لمشاهد (بيانات) صحيحة. إذ لابد من الإشارة إلى أن دقة المعلومة المكتسبة من خلال التمثيل البياني الإحصائي تعتمد بالطبع على دقة البيانات الأصلية ومدى خلوها من الأخطاء، فالرسم البياني يمثل أداة قوية لعرض المعلومات إذا ما تم استخدامه بشكل صحيح واعتمادا على دقة وصحة البيانات سواء من مصدرها أو من خلال عملية الإدخال.

وكثيرا ما نرى العرض البياني الإحصائي يوميا في معظم مجالات الحياة ووسائل الإعلام مثل الصحف والتلفاز وعلى شبكة الانترنت، ففي الاقتصاد مثلا نشاهد الرسومات التي تدل على التغيرات اليومية في تداولات البورصة وحركة الأسواق والأسهم وأسعار صرف العملات وأسعار النفط وغيرها، وفي مجالات الطب قد نرى إحصائية لمنظمة الصحة العالمية ممثلة بيانيا تحذر من خطورة السفر إلى بعض المناطق بسبب انتشار وباء معين، (كما شاهدنا سابقا عندما تفشى وباء السارس وأنفلونزا الطيور وأنفلونزا الخنازير ...)، وحتى في الدراسات الاجتماعية كثيرا ما نلاحظ بعض الرسومات البيانية التي تظهر ارتفاع نسب الذين تأخر زواجهم مثلا في أحد الدول على مر السنوات، ...، وهكذا.

لهذا، فإنه كما انتشرت ثقافة الوجبات السريعة (Fast Food) في كل المجتمعات ومن ضمنها العربية، فإن ثقافة الحصول على المعلومة السريعة (Fast Information) قد بدأت تنتشر بسرعة أيضا خلال السنوات الأخيرة. فالكثيرون، ممن يستخدمون الأساليب الإحصائية، يودون الحصول على المعلومة المفيدة، سواء كانت وصفية أو تحليلية عميقة، بشكل سريع وبعيدا عن التفاصيل المملة التي قد لا تعنيه برأيه. وهذه في الحقيقة هي أحد الملهمات المستحدثة في علم الإحصاء مؤخرا والتي تعول على الباحث الإحصائي في استيعاب كم هائل من البيانات ومعالجتها وتحليلها، باستخدام الأساليب التقليدية والحديثة لإحصاء والقدرات الفائقة المتنامية للحاسوب (من كيان مادي وبرمجيات)، ومن ثمة وصفها وعرض أهم النتائج التي تم التحصل عليها بأفضل طريقة ممكنة بعيدا عن العرض النظري الممل أحيانا بالنسبة لغير المتخصصين[1].

ولابد من الإشارة إلى أن دقة المعلومة المكتسبة من خلال الرسم البياني تعتمد بالطبع على دقة البيانات الأصلية ومدى خلوها من الأخطاء (المقصودة وغير المقصودة). فالرسم البياني يمثل أداة قوية لعرض المعلومات، (وكذلك استكشافها)، إذا ما تم استخدامه بشكل صحيح يتناسب مع طبيعة البيانات ونوعها وأيضا مع نوع الدراسة المطلوبة، على اعتبار أن البيانات المستخدمة دقيقة وصحيحة.

ولكي نتمكن من وصف البيانات بصورة واضحة وموجزة لابد، كما أسلفنا، من استخدام العرض أو الرسم البياني المناسب. وهنالك عدة أنواع من الرسومات البيانية في ساحة الإحصاء والتي تستخدم لتمثيل مفردات البيانات وإيصال المعلومات بصورة سريعة، وسنقوم فيما يلي بتوضيح أشهرها وأكثرها استخداما في عملية التحليل الإحصائي الاستكشافي.

[1] وهذا ببساطة أحد مفاهيم **تنقيب البيانات** الذي أشرنا له في الفصل الأول.

1.3.2 الرسم النقطي (Dot or Point Diagram)

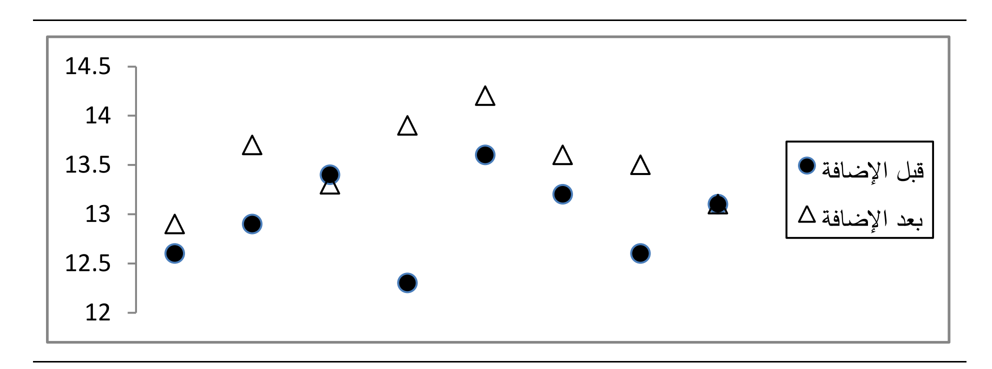
ويعد هذا الرسم من أبسط أنواع الرسومات البيانية على الإطلاق، ويستخدم عادة لتمثيل عدد محدود (ضئيل) من البيانات المفردة الغير مصحوبة بتكرار على خط الأعداد باتجاه واحد، أو المصحوبة بتكرار على محورين. وهو يتيح لنا مراقبة موضع البيانات وانتشارها (تغير قيمها)، وبالتالي تحديد سلوكها، وكذلك يمكن استخدامه للمقارنة بين متغيرين أو أكثر. كما أنه يمكن استخدام الرسم النقطي لتمثيل المتغيرات الوصفية المرتبطة بالتكرارات.

مثال (5.2): أراد أحد المهندسين إجراء دراسة تتعلق بقوة تحمل خليط من المعادن المستخدمة في بناء جدران المنشئات العسكرية. فقام في البداية بعمل نموذج أولي وتم اختباره 8 مرات مع زيادة درجة الحرارة المحيطة بالتدريج. ثم قام هذا المهندس بإضافة معدن جديد للخليط وقام بإعادة الاختبار تحت نفس درجات الحرارة المستخدمة مع الخليط الأول فكانت النتائج، (والتي تمثل مقاومة الخليط بالكيلوجرام)، كما هو موضح في الجدول (8.2).

جدول (8.2): نتائج مقاومة خليط المعادن قبل وبعد إضافة المعدن الجديد.

المتغير X (المقاومة قبل الإضافة)	12.6	12.9	13.4	12.3	13.6	13.2	12.6	13.1
المتغير Y (المقاومة بعد الإضافة)	12.9	13.7	13.3	13.9	14.2	13.6	13.5	13.1

نقوم الآن بتوزيع (تمثيل) قيم كل متغير على خط الأعداد الصحيحة فنحصل ببساطة على الرسم النقطي للبيانات في جدول (8.2) ممثلة في الشكل (1.2) كما يلي.



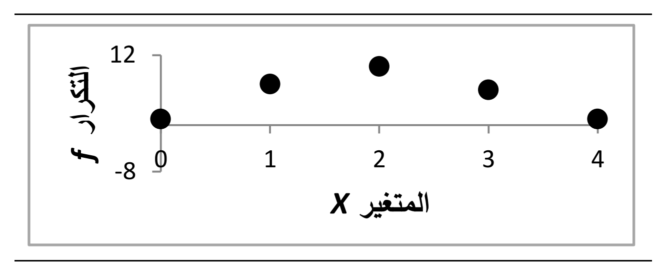
شكل (1.2): الرسم النقطي للبيانات في جدول (8.2).

نلاحظ من الشكل (1.2) ما يلي: أولا بزيادة درجات الحرارة فإن مقاومة الخليط سواء قبل إضافة المعدن الجديد (ويرمز لذلك الدوائر) أو بعد الإضافة (ويرمز لذلك المثلثات) تزداد بدورها في العموم. وثانيا كمقارنة بين المتغيرين (X و Y) فإن إضافة المعدن الجديد قد أدى إلى زيادة مقاومة الخليط بصورة عامة.

ويمكن استخدام الرسم النقطي لتمثيل المتغيرات المصحوبة بتكرارات في مستوين (محورين) كما يوضح المثال التالي:

مثال (6.2): البيانات في المثال (2.2)، والخاصة بأعداد الزوار لمجموعة من المرضى في إحدى المستشفيات، يمكن تمثيلها باستخدام الرسم النقطي كما هو موضح في الشكل (2.2).

ونرى من الشكل السابق أن معظم المرضى قد تلقوا زيارات أعداد الأشخاص فيها تتراوح مابين زائر وثلاث زوار خلال فترة إقامتهم في المستشفى.

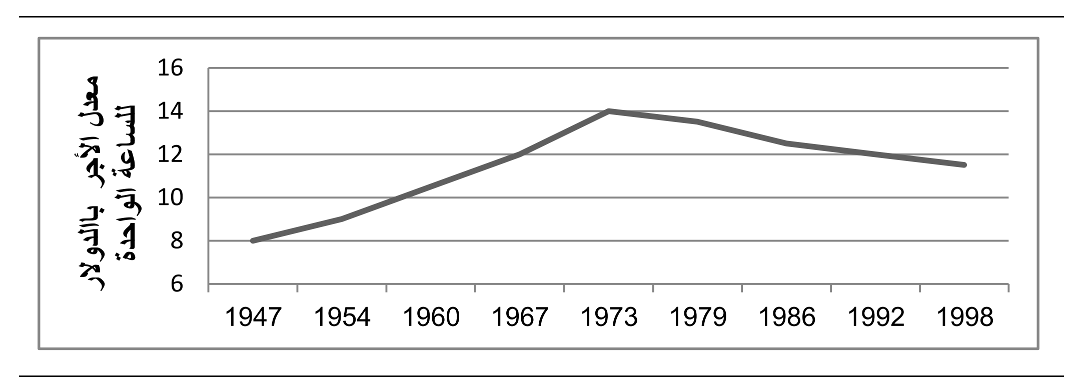


شكل (2.2): الرسم النقطي للبيانات في المثال (2.2).

2.3.2 مخطط الزمن (Time Chart)

وهو نوع من الرسومات البيانية التي تهدف إلى توضيح التغير في مفردات البيانات خلال فترة زمنية معينة (سنوات، أشهر،...، أو حتى مواسم). وهو يشبه الرسم النقطي إلى حد كبير إلا أنه يأخذ بالاعتبار التغير عبر الخط الزمني. ويستخدم هذا النوع من الرسومات في الكثير من المجالات أبرزها عرض التقارير الاقتصادية والصناعية. والتمثيل البياني لمخطط الزمن يتم برصد نقطة لكل قيمة من قيم المتغير مقابل كل وحدة زمنية ثم توصيل هذه النقاط بخطوط مستقيمة.

مثال (7.2): الشكل التالي[1] (شكل (3.2)) يوضح معدل أجور العمال للساعة الواحدة في المجتمع الأمريكي في الفترة ما بين 1947 و 1998 بالدولار الأمريكي.



شكل (3.2): معدل أجور العمال في الولايات المتحدة في الفترة من 1947 إلى 1998

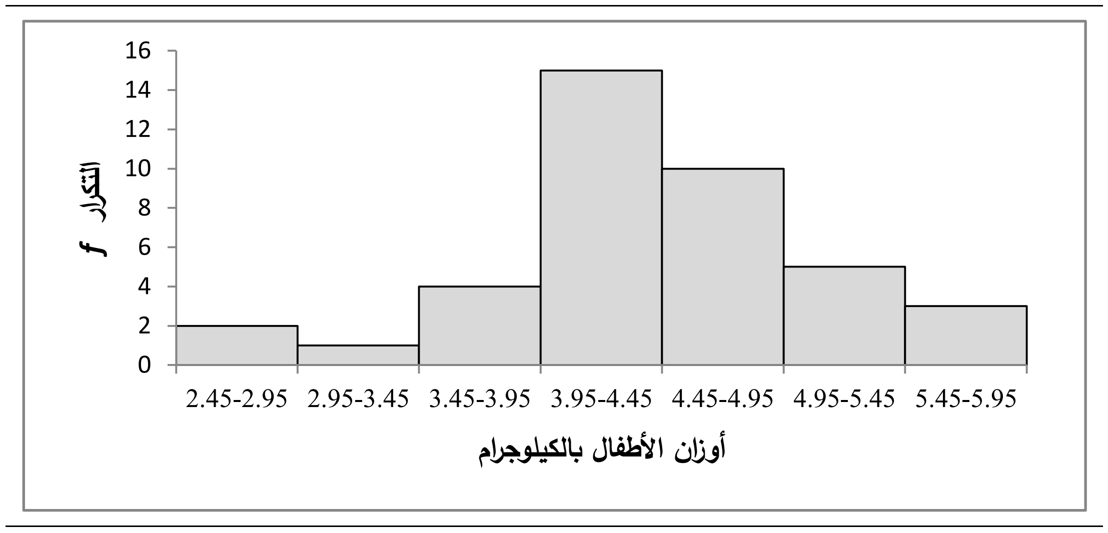
ونرى من الشكل أن معدل أجور العمال تزايد من العام 1947 إلى العام 1973 ثم أخذ بالانخفاض تدريجيا حتى العام 1998 وهذا يعد طبيعيا بسبب التطور الكبير في مجال تقنية آلات المصانع في منتصف السبعينات في الولايات المتحدة مما أدى إلى انكماش دور العامل يدويا في مرحلة الإنتاج.

3.3.2 المدرج التكراري (Histogram)

يعد المدرج التكراري من أشهر وسائل الإيضاح المرئي الذي قد لا تكاد تخلو منه أي دراسة إحصائية استكشافية. ويستخدم لتمثيل البيانات أو المتغيرات الكمية الموجودة في جداول التوزيع التكراري، وذلك عن طريق تمثيل فترات الجداول (المستمرة) كقواعد لأعمدة طولها يمثل التكرار المناظر لكل فترة.

[1] بيانات (رامزي (Ramzy)، 2003).

مثال (8.2): باعتبار الفترات المستمرة والتكرارات المناظرة لها في جدول (6.2)، (مثال (4.2))، نستطيع تكوين قواعد المدرج التكراري بداية من الفترة 2.45–2.95 بحيث يبدأ العامود التالي من الحد الأعلى للفترة السابقة وهكذا حتى آخر عامود، كما هو موضح في الشكل (4.2)، والذي يبين لنا أن أوزان الأطفال المسجلين يتركز معظمها ما بين 3.95 و 4.95 كيلوجرام، وأن الأطفال الذين تقل أوزانهم أو تزيد عن هذه الفترة عددهم قليل وهذا يعد طبيعيا من الناحية الطبية.



شكل (4.2): المدرج التكراري لتوزيع أوزان الأطفال في جدول (6.2).

4.3.2 المضلع التكراري (Frequency Polygon)

إضافة إلى المدرج التكراري، فإن المضلع التكراري يعد أداة مفيدة تستخدم لتمثيل البيانات الكمية بشكل بياني واضح وسهل التفسير. فهو يوضح التغير في قيم التكرارات أيضا ولكن ليس للفترات بحد ذاتها بل لمراكز الفترات. فتكون هذه المراكز ممثلة في المحور الأفقي ويتم رسم نقطة الالتقاء لها مع التكرارات المناظرة في المحور العمودي، كما سنرى في المثال التالي:

مثال (9.2): من البيانات في جدول (6.2)، (مثال (4.2))، نقوم بتمثيل توزيع مراكز الفترات في المحور الأفقي والتكرارات في المحور العمودي فنحصل على مجموعة من النقاط المنتشرة. بعد ذلك يتم توصيل هذه النقاط بخطوط مستقيمة فنحصل على رسم المضلع التكراري كما هو موضح في الشكل (5.2)، والذي نلاحظ منه ما يلي:

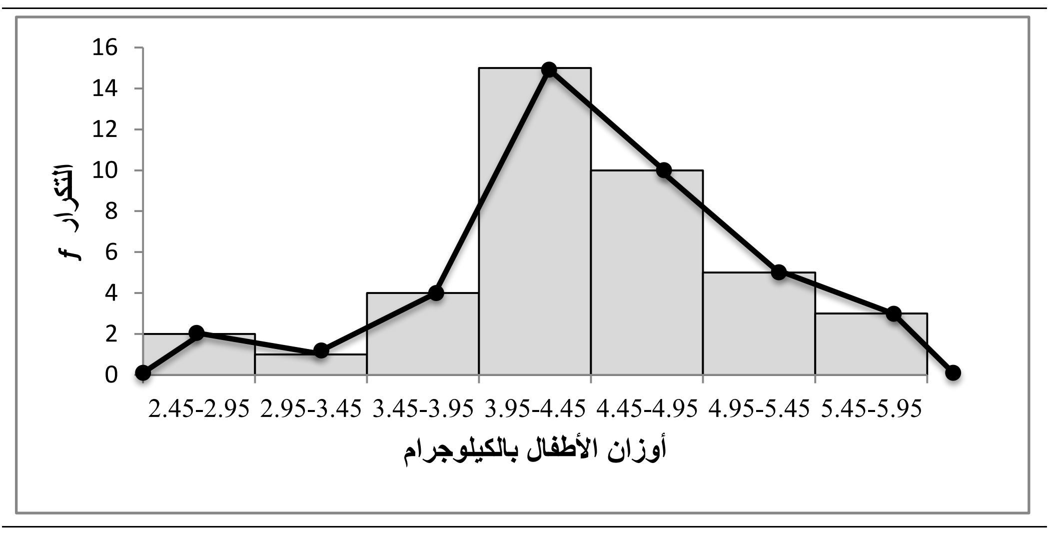
1. استحداث المركزين (2.2) في البداية و(6.2) في النهاية في المحور الأفقي من الشكل، وقد تم ذلك لجعل المضلع التكراري مغلقا من الطرفين. وتم حساب قيمة المركز الأول بطرح قيمة طول الفترة وهي (L= 0.5) من قيمة أول مركز فترة (2.7). وحساب القيمة الأخيرة بإضافة قيمة طول الفترة لقيمة آخر مركز فترة (5.7).



شكل (5.2): المضلع التكراري لتوزيع أوزان الأطفال في جدول (6.2).

2. يمكن استخدام التكرار النسبي لرسم المضلع التكراري بحيث يكون ممثلا في المحور العمودي ويسمى الشكل في هذه الحالة **بالمضلع التكراري النسبي** (Relative Frequency Polygon).

3. يمكن تمثيل كلا من المدرج التكراري والمضلع التكراري في شكل مدمج وذلك عن طريق تحديد نقاط في منتصف الخط العلوي لكل عامود من أعمدة المدرج التكراري ثم وصلها ببعضها البعض بخطوط مستقيمة كما يوضح الشكل (6.2).

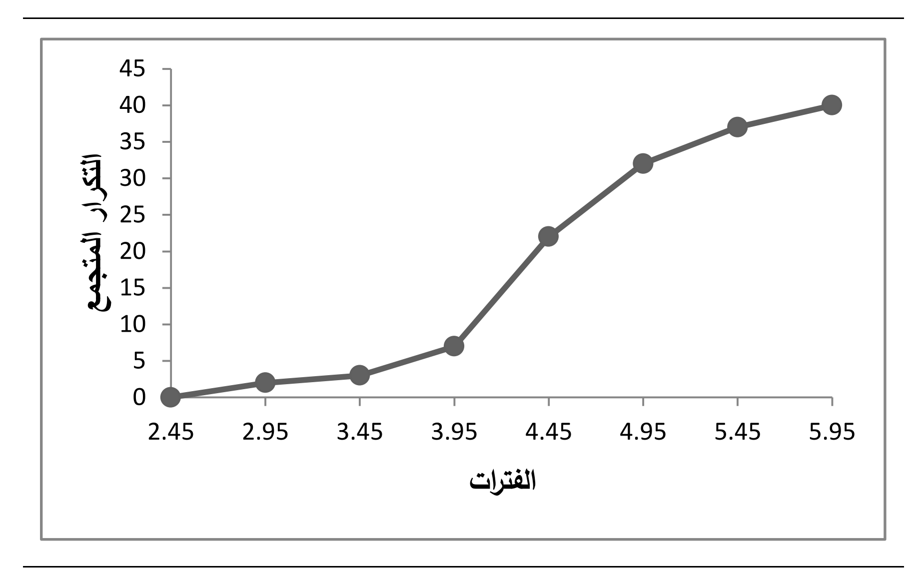


شكل (6.2): رسم المضلع التكراري من خلال المدرج التكراري للبيانات في جدول (6.2).

5.3.2 المضلع التكراري المتجمع (Frequency Ogive)

والذي يعرف بهذا الاسم نسبة إلى شكله المقوس، ويختلف عن المضلع التكراري بأنه يأخذ شكل القوس دائما ويتم رسمه من خلال اعتبار حدود الفترات المستمرة في المحور الأفقي وقيم التكرار المتجمع في المحور العمودي، ثم يتم توصيل النقاط بخطوط مستقيمة فنحصل على الرسم المطلوب.

مثال (10.2): من البيانات في الجدول (7.2)، (مثال (4.2))، نستخدم الفترات المستمرة وقيم التكرار المتجمع الصاعد فنحصل على المضلع التكراري المتجمع المبين في شكل (7.2). والهدف من استخدام المضلع التكراري المتجمع يكون عادة لمراقبة نمط التدرج في زيادة قيم التكرارات (الوثبات) ضمن الفترات المتتالية.



شكل (7.2): المضلع التكراري المتجمع الذي يمثل توزيع أوزان الأطفال في جدول (7.2).

ونلاحظ من الشكل (7.2) أن الزيادة أو الوثبة الرئيسية كانت في الفترة من 3.95 إلى 4.45، أي أن عدد الأطفال الذين تقع أوزانهم في هذه الفترة كان مؤثرا في هذه الظاهرة.

ملاحظات:

1. قياسا على ما قد سبق فإنه إذا ما تم استخدام التكرار المتجمع النسبي في شكل (7.2) فإننا نحصل على **المضلع التكراري المتجمع النسبي** (Relative Frequency Ogive).

2. إذا ما تم توصيل النقاط (إحداثيات محور X و Y) بشكل منحني وليس بخطوط مستقيمة فإن الشكل عندئذ يسمى **المنحنى التكراري**.

6.3.2 الأعمدة البيانية (Bar Charts)

إن رسومات الأعمدة البيانية، من حيث الشكل العام، تشبه رسومات المدرج التكراري. إلا أن الأعمدة البيانية غالبا[1] ما تستخدم لتمثيل وعرض البيانات الوصفية غير الرقمية. وتكون الطريقة بأن يتم توزيع قيم المتغير المختلفة على أحد المحاور، (المحور الأفقي عادة)، وتكرار الحدوث أو النسبة على المحور الآخر.

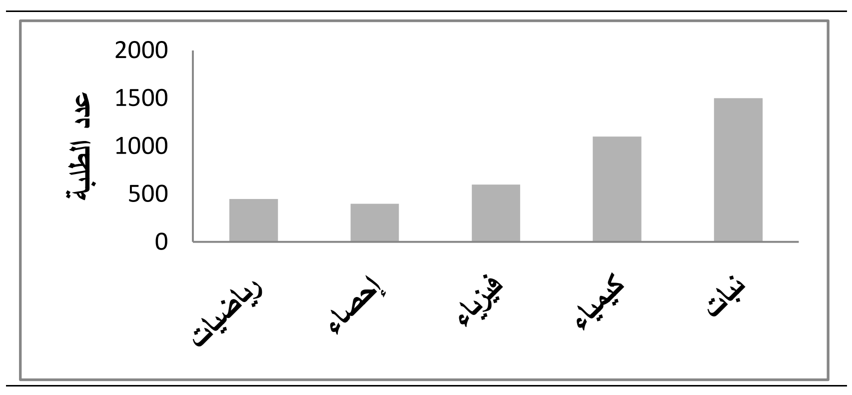
مثال (11.2): البيانات في الجدول التالي، (جدول (9.2))، تمثل أعداد الطلبة الذكور والإناث في خمسة أقسام في كلية العلوم بجامعة بنغازي لعام 2011 . وباستخدام الأعمدة البيانية يمكن "تحويل" محتويات الجدول إلى "صورة" معبرة عن ما تمثله هذه البيانات.

[1] بعض الإحصائيين يستخدمون الأعمدة البيانية لتمثيل البيانات الرقمية عند التعامل مع الفترات غير المستمرة، (ولبول (Walpole)).

جدول (9.2): توزيع أعداد الطلبة الذكور والإناث في أقسام كلية العلوم بجامعة بنغازي.

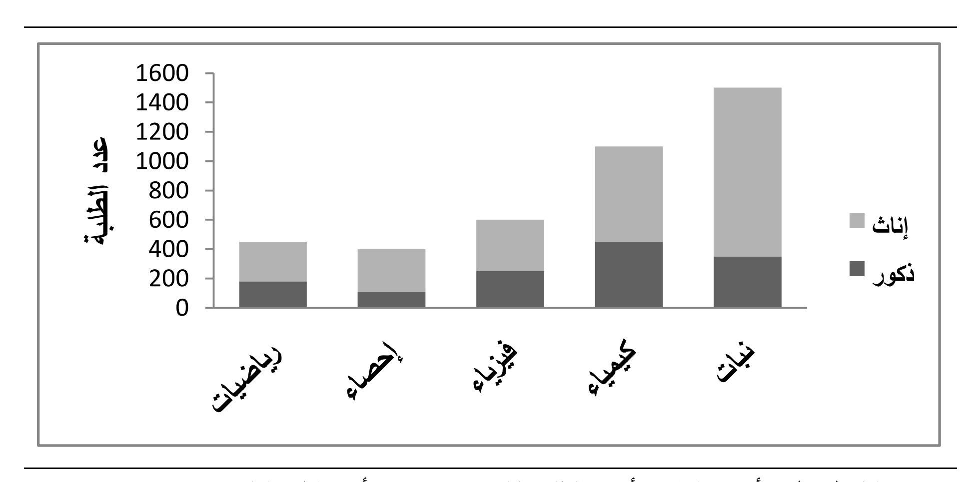
	القسم				
النوع ↓	**رياضيات**	**إحصاء**	**فيزياء**	**كيمياء**	**نبات**
ذكور	180	110	250	450	350
إناث	270	290	350	650	1150
المجموع	450	400	600	1100	1500

ويتضح من شكل (8.2) أن قسمي الكيمياء والنبات هي من أكثر الأقسام اكتظاظا بالطلبة في كلية العلوم، وأن قسم الإحصاء هو أقلها في عدد الطلبة.



شكل (8.2): الأعمدة البيانية لأعداد الطلبة في أقسام كلية العلوم بجامعة بنغازي.

ومن ضمن الاستخدامات الواسعة للأعمدة البيانية إمكانية إدراج تصنيف إضافي للمتغير الأصلي في نفس الرسم. ففي المثال السابق، (مثال (11.2))، يمكن توضيح أو إظهار أعداد الطلبة الذكور والإناث في كل قسم من أقسام الكلية كما هو موضح في شكل (9.2)، والذي يعكس تفوق عدد الإناث في كل الأقسام على عدد الذكور بنسبة كبيرة، وتجدر الإشارة هنا إلى أن هذه الزيادة في عدد الطالبات الجامعيات مقارنة بأعداد الطلبة الذكور هي الصفة السائدة في جامعة بنغازي في العموم.



شكل (9.2): الأعمدة البيانية لأعداد الطلبة الذكور والإناث في أقسام كلية العلوم بجامعة بنغازي.

7.3.2 القطاعات الدائرية (Pie Charts)

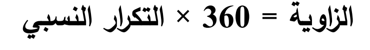
تعد القطاعات الدائرية من الرسومات البيانية الأكثر استخداما لتمثيل البيانات بيانيا نظرا لسهولة استخدامها وبساطة تلخيصها للبيانات. وتستخدم هي الأخرى، في الغالب، لتمثيل البيانات الوصفية. حيث يعبر كل قطاع من قطاعات الدائرة الممثلة لكل مفردات المتغير على نسبة تصنيف معين من التصنيفات الكلية بحيث يكون مجموع نسب هذه التصنيفات هو 100%.

مثال (12.2): البيانات في جدول (10.2) تمثل توزيع أعداد المرضى المصابين بمرض السرطان بحسب مكان الإصابة، وذلك في أحد مراكز علاج السرطان في إحدى المدن الأمريكية.

جدول (10.2): توزيع أعداد مرضى السرطان حسب مكان الإصابة، وحساب التكرار النسبي وحجم الزاوية للمثال (12.2).

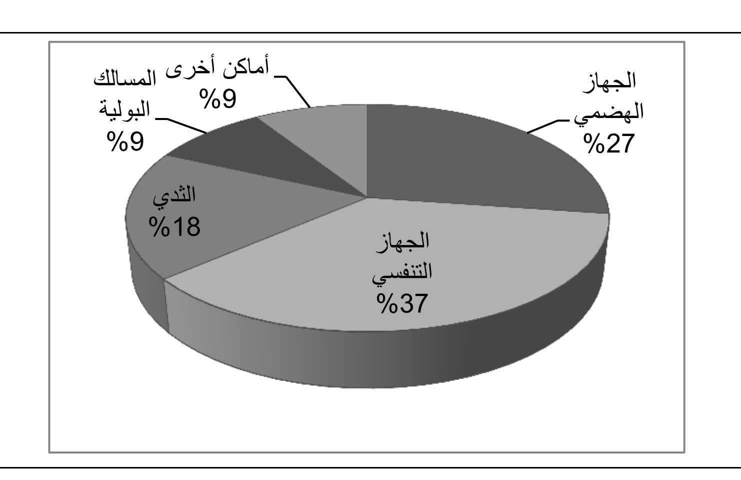
مكان الإصابة	عدد المصابين	التكرار النسبي	حجم الزاوية
الجهاز الهضمي	30	0.27	$360\times0.27 = 97.2^o$
الجهاز التنفسي	40	0.36	$360\times0.36 = 129.6^o$
الثدي	20	0.18	$360\times0.18 = 64.8^o$
المسالك البولية	10	0.09	$360\times0.09 = 32.4^o$
أماكن أخرى	10	0.09	$360\times0.09 = 32.4^o$
المجموع[1]	110	1	360^o

ولرسم القطاع الدائري، يتم حساب التكرار النسبي بقسمة كل مفردة من مفردات المتغير على مجموع المفردات، ثم حساب أحجام الزوايا باستخدام الصيغة التالية:



الزاوية = 360 × التكرار النسبي

كما هو موضح في العامودين الأخيرين في جدول (10.2). ثم يتم رسم هذه الزوايا الواحدة تلو الأخرى داخل الدائرة فنحصل على رسم القطاع الدائري كما هو موضح في شكل (10.2).



شكل (10.2): القطاع الدائري لنسب مرضى السرطان حسب مكان الإصابة للمثال (12.2).

[1] سيلاحظ القارئ هنا أن مجموع التكرار النسبي لا يساوي الواحد الصحيح تماما وذلك بسبب خطأ التدوير (Rounding Error) إلى خانتين والذي قصدنا إبرازه في هذا المثال، وكذلك الأمر مع مجموع الزوايا. وقد تمت الإشارة في موضع آخر من الكتاب إلى أن استخدام الحزم الإحصائية يغني عن أخطاء التدوير الناتجة عن استخدام الآلة الحاسبة التقليدية.

ملاحظة: إضافة إلى ما سبق عرضه من الرسومات البيانية، فإنه يوجد أسلوب رسم بياني إضافي يعرف باسم **شكل الصندوق** (Boxplot or Box and Whisker plots)، والذي على الرغم من أهميته الكبيرة في مجال استكشاف ووصف توزيع البيانات، والمقارنة بين عدة متغيرات من حيث توزيع مفرداتها، إلا أن كثيرا من الباحثين أو مستخدمي الأدوات الاستكشافية الإحصائية يعزُفون أحيانا عن استخدامه رغم بساطته. وسيتم مناقشة شكل الصندوق لاحقا في الفصل القادم حيث أنه يعتمد على حساب بعض مقاييس النزعة المركزية، والتي سيتم عرضها في الجزء التالي، ومقاييس التشتت، والتي سنتناولها بالشرح في الفصل القادم ضمن الجزء الثاني في أساليب التحليل الاستكشافي للبيانات.

4.2 مقاييس النزعة المركزية (Measurements of Central Tendency)

ناقشنا في ما سبق كيف أن الإحصاء الاستكشافي يهدف إلى تنظيم وتلخيص قاعدة البيانات وعرضها بصورة واضحة بغية استخلاص معلومات حول توزيع وسلوك مفردات البيانات ووصف الظاهرة ككل. ومن ضمن الاستراتجيات الإحصائية المستخدمة في الوصف والتحليل الاستكشافي اختزال عدة قيم (مفردات المتغير) في قيمة واحدة معبرة. وهذا بالطبع يعني أنه سيتم "التضحية" بقدر معين من المعلومات الكلية الكامنة في البيانات الأصلية، إلا أن هذه الاستراتجيات إذا ما استخدمت بطريقة سليمة، واعتمادا على صحة ودقة البيانات، فإن القدر المستخلص من المعلومات من خلال تلك المقاييس المختزلة يكون عادة كافيا للتعبير عن الظاهرة.

ولتوضيح هذه النقطة، نأخذ المثال التوضيحي التالي؛ لنفرض أن أحد طلبة الدراسات العليا في أحد الأقسام العلمية قام بإجراء عدة اختبارات في مجموعة من المواد المختلفة خلال العام الدراسي، وعلى افتراض أن الدرجة النهائية لكل اختبار هي 20 درجة، فعندما نقول أن هذا الطالب قد أحرز "معدلا" قدره 14 درجة في المواد ككل فهذا قد لا يعني أنه قد تحصل على 14 درجة "بالضبط" في لأي مادة، ولكن هذا يعني أن درجاته بصورة عامة تميل أو تتمركز أو "تنزع" إلى هذه القيمة (الدرجة). وهذا ما تهدف **مقاييس النزعة المركزية** لإيضاحه ووصفه من خلال توزع البيانات.

إن أي مقياس حسابي يستخدم لتحديد "مركز" مجموعة من المفردات يسمى مقياس نزعة مركزية أو مقياس **موقع مركزي** (Central Location)، ومن أشهر المقاييس المستخدمة لهذا الغرض، كما سنستعرض لاحقا، هي الوسط والوسيط.

وبصورة عامة، فإن كل مجموعة من البيانات، مهما كانت تمثل، تتمتع بخاصيتين أساسيتين هما **قيمة مركزية** (Central value) و**انتشار** (Spread) حول هذه القيمة، وهاتين الخاصيتين هما اللتان يعتمد عليهما استكشاف ووصف البيانات ببساطة شديدة، حيث إن مقاييس النزعة المركزية، والتي سنستعرضها بالتفصيل في هذا الجزء، تعد إحدى المكونات الرئيسية لمفهوم **الإحصاءات الموجزة أو الملخصة** (Summary Statistics)، والتي تعرف[1] بأنها قيم عددية تستخدم لوصف أو استكشاف الصفات المميزة لتوزيع البيانات. أما المكون الثاني للإحصاءات الموجزة فهو مقاييس التشتت أو الانتشار، والذي سيكون موضوع الفصل القادم. إن الإحصاءات

[1] (لا بلانك (La Blanc)، 2004).

الموجزة تكون عادة أداة مفيدة لتوضيح مركز البيانات، ويقصد بالمركز القيمة التي تتوسط مجموعة البيانات، ومن المفترض، (رغم عدم تحقق ذلك في بعض الحالات)، أن تتمركز أو تتوزع معظم قيم البيانات حول هذا المركز، ولهذا يقال أن القيم تنزع عادة إلى مركزها.

وسنقوم فيما يلي بعرض أهم المقاييس الخاصة بالنزعة المركزية وتوضيح كيفية حسابها باستخدام البيانات المفردة (غير المبوبة)، والبيانات المبوبة (ذات المشاهدات المتكررة وذات التوزيعات التكرارية). وسنبدأ بتوضيح كل مقياس من هذه المقاييس باستخدام البيانات المفردة في البداية.

1.4.2 الوسط (Mean)

يعتبر الوسط أو **المتوسط** من أهم وأشهر مقاييس النزعة المركزية وأكثرها استخداما على الإطلاق، ويعرف أيضا باسم **المعدل** (Average)، ويطلق عادة مصطلح الوسط على **الوسط الحسابي** بالتحديد، نظرا لوجود عدة صيغ أخرى تستخدم لحساب الأوساط كما سنرى لاحقا.

ولتوضيح مفهوم الوسط الحسابي[1] كمقياس وصفي نطرح التساؤل التالي بصورة مجردة:

كيف يمكن أن نصِف مجموعة من البيانات؟ . للإجابة عن هذا التساؤل نستخدم المثال التوضيحي التالي:

ظهرت نتائج مجموعة من الطلاب في مقرر مبادئ الإحصاء في نهاية الفصل الدراسي كالتالي:

53	71	65	40	35	55	62	75	50	79	45	30

وكان المطلوب هو وصف المستوى الدراسي للطلبة في هذا المقرر. في الواقع، إن أبسط طريقة لوصف هذه البيانات يمكن أن تكون بإيجاد قيمة واحدة، (درجة في هذه الحالة)، بحيث تعبّر عن كل درجات الطلبة وتصف مستواهم الدراسي. هذه القيمة يمكن أن تكون الوسط الحسابي أو المعدل، (والذي ستساوي قيمته 55 درجة، حيث سيتم حسابه باستخدام التعريف القادم)، هذا الوسط يعطينا الانطباع بأن المستوى الدراسي لهؤلاء الطلبة متدني جدا، وهذا قد يدفع المحاضر للبحث عن السبب أو الأسباب الكامنة وراء هذه النتيجة.

وهكذا يمكننا القول بأن الوسط الحسابي هو قيمة مركزية تتجمع حولها مجموعة من القيم، ويمكن من خلالها "الحكم" على كل المجموعة. وسنلاحظ أنه لو حلت قيمة الوسط محل كل مفردة من مفردات البيانات لكان مجموع هذه القيم مساويا لمجموع القيم الأصلية[2].

تعريف (1.2): **الوسط الحسابي** (Arithmetic Mean, A.M): إذا كانت $x_1, x_2, \dots, x_N$ هي مفردات بيانات منتهية حجمها N، فإن الوسط الحسابي لها، والذي يرمز له بالرمز A.M أو $\bar{X}$، يعرف بالصيغة:

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$$

[1] تجدر الإشارة هنا إلى أن الوسط الحسابي يستخدم لوصف البيانات الكمية لا النوعية.

[2] أي أنه في مثالنا التوضيحي لو استبدلنا كل درجات الطلبة بالدرجة 55 فإن المجموع سيساوي مجموع القيم الأصلية وهو 660 .

أي أن الوسط الحسابي لمجموعة من المفردات هو ببساطة مجموع قيمها مقسوما على عددها.

مثال (13.2): أوجد الوسط الحسابي للبيانات التالية، (جدول (11.2))، والتي تمثل أعداد المرضى المصابين بمرض فقر الدم الحاد، والموجودين في كل مستشفيات إحدى المدن الصغيرة وعددها خمسة مستشفيات.

جدول (11.2): توزيع أعداد المرضى في خمسة مستشفيات في إحدى المدن.

المستشفى	1	2	3	4	5
عدد المرضى	30	50	60	40	60

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{30+50+60+40+60}{5} = 48$$

أي أن متوسط أو معدل عدد مرضى فقر الدم الحاد في هذه المدينة هو 48 مريضا لكل مستشفى، وهذا قد يعد مؤشرا هاما على ارتفاع عدد المصابين بهذا المرض، إذا ما كان عدد السكان في هذه المدينة صغير نسبيا، مما يستدعي المزيد من البحث للوقوف على الأسباب المؤدية لذلك.

1.1.4.2 **خواص الوسط الحسابي** (Particularities of Arithmetic Mean)

يتمتع الوسط الحسابي بعدة خواص منها الجيد ومنها غير ذلك كما سنرى في التوضيح التالي:

1. يعتمد الوسط الحسابي على قيمة كل مفردة (x_i , i=1, 2, ..., N) من مفردات البيانات.
2. مجموع انحرافات قيم المفردات عن وسطها الحسابي يساوي دائما الصفر، أي أن:

$$\sum_{i=1}^{N} (x_i - \bar{X}) = 0$$

مثال(14.2): للبيانات 1، 7، 4، 3، 5 يكون $\bar{X} = \frac{\sum_{i=1}^{5} x_i}{5} = 4$ ، وبالتالي فإن انحرافات القيم، (كل قيمة)، عن وسطها الحسابي 4 هو:

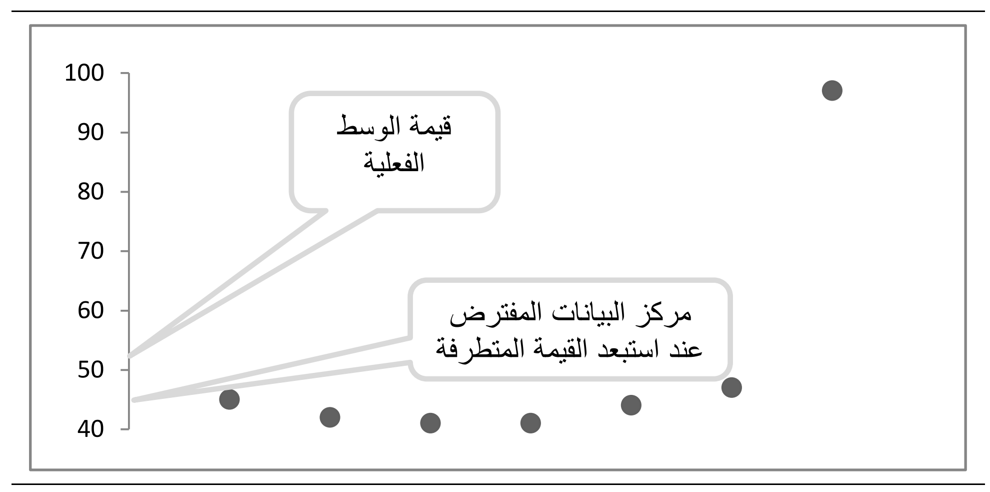
المفردة	1	7	4	3	5
$x_i - \bar{X}$	−3	3	0	−1	1

وبالتالي يكون:

$$\sum_{i=1}^{5} (x_i - \bar{X}) = (-3) + 3 + 0 + (-1) = 0$$

3. من عيوب الوسط الحسابي أنه شديد التأثر بوجود القيم المتطرفة في البيانات، سواء كانت هذه القيم المتطرفة كبيرة جدا، أو صغيرة جدا مقارنة بقيم البيانات ككل.

مثال (15.2): إذا كانت أسعار النفط الخام (بالدولار لكل برميل) لسبعة أشهر هي 45، 42، 41، 41، 44، 47، 97. فإن معدل أسعار النفط سيكون $\bar{X} = 51$ دولار للبرميل، وهذا يعطينا معلومة غير دقيقة عن مستوى أسعار النفط خلال هذه الفترة لأن القيمة المتطرفة 97 في البيانات قد أثرت على قيمة الوسط عن طريق "جذبها" للمركز إلى اليمين، كما يوضح لنا الشكل (11.2).



شكل (11.2): الشكل النقطي لأسعار النفط (دولار/برميل) لسبعة أشهر.

4. مجموع مربعات انحرافات قيم البيانات عن وسطها الحسابي يكون دائما أقل ما يمكن فيما إذا تم استخدام أي قيمة أخرى بديلة. بمعنى أنه لأي مفردات $x_1, x_2, \dots, x_N$ يكون $\sum_{i=1}^{N}(x_i-\bar{X})^2 < \sum_{i=1}^{N}(x_i-X_o)$ ، حيث X_o هي أي قيمة أخرى بحيث $X_o \neq \bar{X}$.

مثال (16.2): للبيانات التالية 2، 4، 6، 8 يكون الوسط الحسابي $\bar{X}=5$، وبافتراض أي قيمتين 3 ، X_o= 6، فإن مجموع مربع انحرافات قيم البيانات عن وسطها الحسابي وعن هاتين القيمتين يكون:

المفردة	2	4	6	8	المجموع
$(x_i-\bar{X})^2$	9	1	1	9	20
$(x_i-6)^2$	16	4	0	4	24
$(x_i-3)^2$	1	1	9	25	36

مما يؤكد الخاصية السابقة[1].

5. إذا كانت $x_1, x_2, \dots, x_N$ هي مفردات بيانات حجمها N، وكان a هو أي عدد حقيقي ثابت، فإن:

أ) $\frac{\sum_{i=1}^{N}(x_i\pm a)}{N}=\bar{X}\pm a$ ، أي أن الوسط الحسابي لمفردات تم إضافة أو طرح أي عدد ثابت منها يساوي الوسط الحسابي للمفردات الأصلية مضافا أو مطروحا منه هذا العدد. وهذا ما يعرف بالتغير الموضعي في الوسط.

الإثبات: لدينا $\frac{\sum_{i=1}^{N}(x_i\pm a)}{N}=\frac{\sum_{i=1}^{N}x_i}{N}\pm\frac{\sum_{i=1}^{N}a}{N}=\bar{X}\pm\frac{Na}{N}=\bar{X}\pm a$

ب) $\frac{\sum_{i=1}^{N}ax_i}{N}=a\bar{X}$ ، بمعنى أن الوسط الحسابي لمفردات تم ضرب قيمها بعدد ثابت يساوي الوسط الحسابي للمفردات الأصلية مضروبا بهذا العدد. وهذا يعرف بالتغير العاملي في الوسط.

الإثبات: لدينا

$$\frac{\sum_{i=1}^{N}ax_i}{N}=\frac{ax_1+ax_2+\dots+ax_N}{N}=\frac{a(x_1+x_2+\dots+x_N)}{N}=\frac{a\sum_{i=1}^{N}x_i}{N}=a\bar{X}$$

[1] المثال (16.2) يعتبر تطبيقا رقميا لتوضيح الخاصية 4 وليس إثباتا نظريا.

2.1.4.2 بعض الأوساط الأخرى (Some Other Means)

وتوجد بعض مقاييس النزعة المركزية والتي تنتمي لعائلة الأوساط نذكر منها ما يلي:

1. **الوسط الحسابي الموزون** (Weighted Arithmetic Mean):

في بعض الدراسات الخاصة، قد نكون بحاجة لإعطاء أهمية (وزن) لبعض المفردات في مجموعة البيانات أكثر من باقي المفردات نظرا لطبيعة الدراسة. كما هو الحال مثلا عند حساب معدل درجات مجموعة من الأشخاص في اختبار ما علما بأن مستوى الذكاء أو المستوى التعليمي لهم متفاوت جدا، فعندئذ يجب أن نأخذ بالاعتبار وزن كل درجة (وهو ما يمثله مستوى الذكاء أو المستوى التعليمي ممثلا بوحدة قياس معينة[1])، بدلا من حساب الوسط بالطريقة التقليدية.

تعريف (2.2): **الوسط الحسابي الموزون:** إذا كانت $x_1, x_2, \ldots, x_N$ هي مفردات مجموعة من البيانات، وكانت $w_1, w_2, \ldots, w_N$ هي الأوزان التي تم تعيينها لهذه المفردات على الترتيب، فإن الوسط الموزون $\bar{X}_w$ يمكن حسابه بالصيغة:

$$\bar{X}_w = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}$$

مثال (17.2): إحدى شركات السيارات تقوم ببيع ثلاثة أنواع من السيارات بالتقسيط بعد إضافة نسبة محددة إلى سعر السيارة الأصلي كما هو موضح في جدول (12.2)، فما هو معدل (متوسط) سعر بيع السيارات في هذه الشركة؟ (مع الأخذ بالاعتبار قيمة النسبة المضافة للسعر).

جدول (12.2): أنواع وأسعار السيارات بالدولار في المثال (17.2).

نوع السيارة	A	B	C
سعر السيارة	5000	6300	4500
النسبة المضافة للسعر	%10.5	%10.8	%11

الحل:

نستخدم الوسط الموزون حيث تمثل الأوزان هنا الأهمية الاعتبارية لكل نوع من أنواع السيارات $\bar{X}_w = \frac{(5000\times10.5)+(6300\times10.8)+(4500\times11)}{10.5+10.8+11} = 5264.40$ دولار.

ملاحظة: لاحظ أن قيمة الوسط الحسابي الموزون تساوي قيمة الوسط الحسابي التقليدي إذا ما كانت الأوزان كلها متساوية $w_1 = w_2 = \ldots = w_N = w$.

2. **الوسط المُجمع أو المشترك** (Combined Mean):

إن بعض قواعد البيانات قد تحتوي على بعض المتغيرات المتجانسة التي تم تسجيلها باستخدام نفس المقياس ولكن في أماكن أو أوضاع أو أوقات مختلفة. في هذه الحالة، قد يكون من المفيد حساب الوسط الحسابي لكل متغير على حده، (بحسب المكان أو الزمان الخاص به)، ثم حساب وسط عام لهذه الأوساط وهذا ما يعرف بالوسط المُجمع.

---

[1] يمكن الاعتماد على وحدة مقياس الذكاء الاعتيادية (IQ) أو اعتبار درجات معروفة مسبقا لدى المتخصص (مثلا؛ 1، 2، ...، أو 2، 4، ...) بحيث تمثل أوزان تزداد بزيادة تأثير المتغير المصاحب، وهو مستوى الذكاء أو المستوى التعليمي.

تعريف (3.2): **الوسط الحسابي المُجمع:** إذا كان لدينا k مجموعة من البيانات، (أو المتغيرات ضمن البيانات)، أحجامها $N_1, N_2, \ldots, N_k$ وكانت لها الأوساط الحسابية $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k$ على الترتيب، عندئذ يعرف الوسط الحسابي المُجمع $\bar{X}_c$ بالصيغة:

$$\bar{X}_c = \frac{\sum_{i=1}^{k} N_i \bar{X}_i}{\sum_{i=1}^{k} N_i}$$

مثال (18.2): تم تدريس مقرر اللغة العربية ضمن 3 مجموعات حيث كان عدد الطلبة في المجموعات 28، 32، و35 طالبا. وفي نهاية العام الدراسي، كانت معدلات درجات الطلبة في المجموعات الثلاث في الامتحان النهائي هي 83، 70، و 76 درجة على الترتيب. أوجد الوسط الحسابي المجمع (العام) لجميع الطلبة في مقرر اللغة العربية.

الحل:

$$\bar{X}_c = \frac{(28\times83)+(32\times70)+(35\times76)}{28+32+35} = 76.04 \text{ درجة}$$

ملاحظة: لاحظ التشابه في الشكل العام لكلا من صيغتي الوسط الموزون والوسط المجمع، إلا أن الأول يعتبر "دالة" في قيمة الوزن والمفردة بحد ذاتها أما الثاني فهو دالة في عدد المفردات وأوساطها.

3. **الوسط الهندسي** (Geometric Mean, G.M):

يستخدم الوسط الهندسي عادة لحساب المعدل للبيانات التي تتابع فيها قيم المفردات بصورة شبه ثابتة، أي تكون النسبة بين قيم المفردات المتتابعة ثابتة تقريبا، كما هو الحال أحيانا في بيانات المؤشرات الاقتصادية وتعدادات السكان عبر السنوات المتلاحقة، وغيرها. ويتميز الوسط الهندسي بأنه لا يتأثر بالقيم المتطرفة كما هو الحال مع الوسط الحسابي.

تعريف (4.2): **الوسط الهندسي:** يعرف الوسط الهندسي G.M لمجموعة من البيانات مكونة من N مفردة $x_1, x_2, \ldots, x_N$ بأنه الجذر N لحاصل ضربها بعضها ببعض، بمعنى

$$\text{G.M} = \sqrt[N]{x_1 x_2 \ldots x_N}$$

ملاحظة: يمكننا كتابة

$$log(\text{G.M}) = \frac{1}{N} log(x_1 x_2 \ldots x_N)$$
$$= \frac{1}{N}\sum_{i=1}^{N} log(x_i)$$

أي أن لوغاريتم الوسط الهندسي لـ N مفردة موجبة يساوي الوسط الحسابي للوغاريتم هذه المفردات. وبالتالي يكون

$$\text{G.M} = Anti\ (log\text{G.M})$$
$$= 10^{\log(\text{G.M})}$$

مثال (19.2): أوجد الوسط الهندسي للبيانات التالية 12، 10، 7، 6، 6، 5، 3 .

الحل:

$$G.M = \sqrt[7]{3 \times 5 \times 6 \times 6 \times 7 \times 10 \times 12} = \sqrt[7]{453600}$$

أو

$$log(G.M) = \frac{1}{7}\log(453600) = 0.8081$$

$$G.M = 10^{(0.8081)} = 6.43$$

4. **الوسط التوافقي** (Harmonic Mean, H.M):

يعد استخدام الوسط التوافقي محدودا بالمقارنة مع الأوساط الأخرى نظرا لأنه يختص عادة بالتعامل مع نوعية محددة من الحالات أو الظواهر التي تتطلب حساب المعدل لمتغيرات تتغير قيم مفرداتها عند ثبات عامل ملازم لها. مثال على ذلك، حساب معدل سرعات مختلفة لسيارات قيد الاختبار مع تثبيت المسافة المقطوعة، أو حساب متوسط تكلفة مجموعة من السلع التجارية تم شراؤها في أماكن مختلفة باستخدام نفس المبلغ المخصص للشراء، وهكذا.

ويعاب على الوسط التوافقي أنه يتأثر دائما بقيم المفردات الصغيرة لأنه يتعامل، حسابيا، مع مقلوب مفردات البيانات، لذلك يكون عادة مناسبا للتطبيق مع البيانات التي تحوي قيما كبيرة.

تعريف (5.2): **الوسط التوافقي:** يعرف الوسط التوافقي H.M للمفردات $x_1, x_2, \dots, x_N$ بأنه مقلوب الوسط الحسابي لمقلوبات المفردات الأصلية، بمعنى

$$H.M = \frac{1}{\frac{\sum_{i=1}^{N}\frac{1}{x_i}}{N}} = \frac{N}{\sum_{i=1}^{N}\frac{1}{x_i}}$$

مثال (20.2): أوجد الوسط التوافقي للمفردات 80، 40، 20 .

الحل:

$$H.M = \frac{3}{\frac{1}{20}+\frac{1}{40}+\frac{1}{80}} = 34.29$$

ملاحظة: توجد علاقة تربط كلا من الأوساط الثلاثة، الحسابي، الهندسي، والتوافقي كالتالي

$$A.M \geq G.M \geq H.M$$

ويمكن التحقق من تلك العلاقة حسابيا من خلال المثال القادم.

مثال (21.2): أوجد كلا من الوسط الحسابي و الوسط الهندسي و الوسط التوافقي للمفردات التالية 10، 7، 1، 5، 2 ، ثم قارن بين قيم الأوساط الثلاثة.

الحل:

$$A.M = \bar{X} = 5 \; > \; G.M = 3.71 \; > \; H.M = 2.57$$

5. **الوسط الحسابي الفرضي** (Guessed Arithmetic Mean):

توجد طريقة مختصرة لحساب الوسط الحسابي، (تستخدم أحيانا عند التعامل مع قواعد البيانات الضخمة[1])، وهي أن يتم افتراض قيمة مبدئية للوسط، ولتكن A مثلا، ثم يتم إيجاد انحرافات قيم المفردات عن هذا الوسط الفرضي A ومن ثمة حساب الوسط الحسابي الجديد بالصيغة

$$\bar{X}_g = \mathrm{A} + \frac{\sum_{i=1}^{N} d_i}{N}, \quad d_i = x_i - \mathrm{A}\,, i = 1,2, \dots, N$$

بمعنى أن d_i تمثل انحرافات قيم المفردات عن وسطها الفرضي A، وسنقوم بتطبيق أسلوب الوسط الحسابي الفرضي في حالة البيانات المبوبة لاحقا.

2.4.2 الوسيط (Median)

يعتبر مقياس الوسيط ثاني أشهر مقاييس النزعة المركزية استخداما وأهمية بعد الوسط الحسابي، وكما تدل التسمية، فإن الوسيط هي تلك القيمة التي "تتوسط" توزيع مفردات البيانات بحيث يكون عدد المفردات قبل قيمة الوسيط مساويا لعدد المفردات بعد قيمة الوسيط. ومن عيوب الوسيط أنه لا يتناول في (صيغة) حسابه كل قيم المفردات، كما سنرى من تعريفه، إلا أن هذا العيب يعد في نفس الوقت ميزة يتفوق بها مقياس الوسيط على العيب الأساسي في الوسط الحسابي، ألا وهو التأثر بالقيم المتطرفة الكبيرة أو الصغيرة. فالوسيط "ينتقي" القيمة التي تكون في مركز البيانات مع إهمال الأطراف.

تعريف (6.2): **الوسيط:** وسيط أي مجموعة من المفردات مرتبة، تصاعديا (أو تنازليا)، هي تلك القيمة التي تقع في منتصف المفردات عندما يكون عدد المفردات فرديا، أو هي قيمة الوسط الحسابي للقيمتين اللتين تتوسطان المفردات عندما يكون عدد المفردات زوجيا. بمعنى:

إذا كانت $x_1, x_2, \dots, x_N$ هي أي مجموعة من المفردات فإن الوسيط، ويرمز له بالرمز $\tilde{X}$ ، يتم حسابه، بعد ترتيب المفردات تصاعديا (أو تنازليا)، بالصورة:

$$\tilde{X} = \begin{cases} \text{إذا كان } N \text{ فرديا فإن الوسيط هو القيمة التي ترتيبها } \frac{N+1}{2} \\ \text{إذا كان } N \text{ زوجيا فإن الوسيط يساوي الوسط الحسابي للقيمتين التي ترتيبهما } \frac{N}{2} \text{ و } \frac{N}{2}+1 \end{cases}$$

مثال (22.2): أوجد الوسيط للمتغيرين X_1 و X_2 المعطاة قيمهما في جدول (13.2).

جدول (13.2): قيم مفردات المتغيرات في المثال (22.2).

X_1	3	4	4	5	6	8	8	8	10
X_2	5	5	7	9	11	12	15	18	

الحل:

على اعتبار أن قيم المفردات للمتغيرين هي مرتبة تصاعديا مسبقا، فإن الوسيط للمتغير X_1، وعدد مفرداته فردي، هو تلك القيمة التي ترتيبها $5 = \frac{9+1}{2} = \frac{N+1}{2}$ وهي $\tilde{X}_1 = 6$.

[1] في الوقت الحالي، ومع وجود البرامج التي تستطيع التعامل مع أحجام ضخمة جدا من البيانات، أصبح حساب الوسط الحسابي، بالصيغة التقليدية، ممكنا لأي قاعدة بيانات مهما كان حجمها.

والوسيط للمتغير X_2، (حيث أن عدد مفرداته زوجي)، هو الوسط الحسابي للقيمتين $4 = \frac{8}{2} = \frac{N}{2}$ و $5= 1 + \frac{8}{2} = 1 + \frac{N}{2}$ وهو $\tilde{X}_2 = \frac{9+11}{2} = 10$.

3.4.2 المنوال (Mode)

المنوال هو أبسط مقاييس النزعة المركزية مفهوما وحسابا، إلا أن ذلك لا ينقص من أهميته كمؤشر إحصائي هام في منظومة الإحصاء الاستكشافي، وخاصة عند التعامل مع قواعد البيانات الكبيرة، أو مع البيانات التي تتضمن متغيرات وصفية.

تعريف (7.2): **المنوال:** المنوال لمجموعة من المفردات هو القيمة الأكثر ظهورا أو تكرارا. ويرمز له بالرمز M_o أو $\hat{X}$.

وبحسب طبيعة التغير في مفردات البيانات فإن المتغير الذي لا تتكرر قيمه أكثر من مرة لا يكون له منوال، أو قد يكون له منوال واحد (Unimodal)، أو يكون له منوالين ويسمى ثنائي المنوال (Bimodal) إذا كان ضمن المتغير قيمتين لهما أعلى تكرار، ... ، وهكذا يمكن أن يكون المنوال متعددا (Multimodal).

مثال (23.2): أوجد المنوال للمتغيرات في الجدول المرفق، (جدول (14.2)).

جدول (14.2): المتغيرات الخاصة بالمثال (23.2).

X_1	3	5	8	10	12	16	15				
X_2	5	2	2	7	9	12	9	10	9	10	11
X_3	4	3	2	4	4	5	5	7	9	7	7

الحل:

المتغير X_1 ليس له منوال لعدم وجود تكرار لأي قيمة من قيم المفردات. والمتغير X_2 منواله هو $\hat{X}_2 = 9$. أما المتغير X_3 فهو ثنائي المنوال $\hat{X}_3 = 4, 7$.

ولتوضيح مفهوم المنوال بصورة تطبيقية، افترض أن استفتاءا أجري لمعرفة ما هي الرياضة الشخصية المفضلة لدى الناس في مدينة معينة، وكانت الإجابة الأكثر تكرارا (المنوال) هي رياضة المشي، فهذا يعني أن المشي هو النمط السائد ضمن الرياضات الشخصية المختلفة بين الناس في هذه المدينة.

ملاحظة: عندما يكون توزيع البيانات متماثلا[1] تقريبا ولها منوال واحد فقط فإن العلاقة بين المقاييس الثلاثة؛ الوسط والوسيط والمنوال تكون، تقريبا، كالتالي:

$$\bar{X} - \hat{X} = 3(\bar{X} - \tilde{X})$$

وسيتم تقديم مثال على هذه العلاقة عند حساب مقاييس النزعة المركزية لبيانات التوزيعات التكرارية لاحقا.

[1] سيتم التطرق لمفهوم التماثل في توزيع البيانات لاحقا في هذا الفصل.

4.4.2 الربيعات، العشيرات، والمئينات (Quartiles, Deciles, and Percentiles)

لاحظنا سابقا كيف أن الوسيط يمثل القيمة التي تتوسط أو تقسم البيانات إلى جزئين متساويين. يمكننا من هذا المنطلق حساب مقاييس أخرى تقوم بتقسيم مجموعة البيانات إلى أجزاء محددة؛ مثل أربع أجزاء، أو عشرة أجزاء، أو مائة جزء، هذه المقاييس تعرف بصورة عامة باسم **التجزيئات** (Fractiles or Quantiles) . وهذه المقاييس تساعدنا أحيانا في التعرف على طبيعة "تكتل" أو تمركز قيم البيانات وإذا ما كانت هذه القيم تتركز أو تنحسر في منطقة معينة ضمن توزيع البيانات وخاصة عند تعذر مراقبة البيانات بشكل بياني[1].

وسنقوم فيما يلي بتعريف كل مقياس من هذه المقاييس على أن يتم تناول تطبيقاتها على بيانات التوزيعات التكرارية لاحقا.

تعريف (8.2): **الربيعات:** الربيعات هي تلك القيم التي تقسم مجموعة المفردات، بعد ترتيبها تصاعديا، إلى أربعة أجزاء متساوية. ويرمز لهذه القيم بالرموز Q_1 (ويعرف بالربيع الأول أو الأدنى)، Q_2 (ويعرف بالربيع الثاني)، و Q_3 (ويعرف بالربيع الثالث أو الأعلى)، بحيث يكون 25% من المفردات أقل من Q_1، و 50% من المفردات أقل من Q_2، و 75% من المفردات أقل من Q_3.

تعريف (9.2): **العشيرات:** العشيرات، شأنها كشأن الربيعات تقسم البيانات بعد ترتيبها تصاعديا، ولكن إلى عشرة أجزاء متساوية. ويرمز للعشيرات، (وعددها تسعة)، بالرموز D_1، D_2، ... ، D_9 بحيث يكون 10% من المفردات أقل من D_1، و 20% من المفردات D_2، ... ، و 90% من المفردات أقل من D_9.

تعريف (10.2): **المئينات:** المئينات هي تلك القيم التي تقسم البيانات، بعد ترتيبها تصاعديا، إلى مائة جزء متساوي. هذه القيم، (وعددها تسع وتسعون)، يعبر عنها بالرموز P_1، P_2، P_3، ... ، P_{99} بحيث يكون 1% من المفردات أقل من P_1، و 2% من المفردات أقل من P_2، ... ، 99% من المفردات أقل من P_{99}.

ملاحظة: يمكن ملاحظة أن قيمة الربيع الثاني وقيمة العشير الخامس وقيمة المئين الخمسون كلها متساوية وتساوي قيمة الوسيط، أي $Q_2 = D_5 = P_{50} = \tilde{X}$ ، لأنها كلها تقسم المفردات إلى جزئين متساويين.

5.4.2 مقاييس النزعة المركزية لبيانات الجداول التكرارية
(Measurements of Central Tendency for Tabulated Data)

سنقوم في ما يلي بتعريف الصيغ الخاصة بحساب مقاييس النزعة المركزية للبيانات ذات المشاهدات المتكررة وبيانات التوزيعات التكرارية:

تعريف (11.2): **مقاييس النزعة المركزية لبيانات الجداول التكرارية** (البيانات المبوبة): إذا كانت $x_1, x_2,..., x_k$ هي مفردات (أو قيم مراكز فترات لتوزيع تكراري) مناظرة للتكرارات $f_1, f_2, ... , f_k$ على الترتيب[2]، حيث

[1] يُقصد هنا متابعة الباحث للبيانات دون اللجوء لاستخدام الرسومات البيانية.

[2] لاحظ أن k تساوي عدد الخلايا في البيانات ذات التكرارات المشاهدة، وتساوي عدد الفترات في بيانات الجداول التكرارية.

$\sum_{i=1}^{k} f_i = N$ ، فإن مقاييس النزعة المركزية، (التي تم استعراضها سابقا)، يتم حسابها بالصيغ الموضحة أدناه:

1. الوسط الحسابي: ويعطى بالصيغة:

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$$

مثال (24.2): أوجد الوسط الحسابي للبيانات[1] في الجدول التالي، (الجدول (15.2)).

جدول (15.2): بيانات ذات مشاهدات متكررة والخاصة بالمثال (24.2).

X	2	6	8	5	المجموع
f	1	4	2	3	10

الحل:

$$\bar{X} = \frac{(3 \times 5) + (2 \times 8) + (6 \times 4) + (2 \times 1)}{3 + 2 + 4 + 1} = 5.7$$

2. الوسط الهندسي: ويتم حسابه بالصيغة:

$$\mathrm{G.M} = Anti\ (\log(\mathrm{G.M}))$$

حيث

$$\log(\mathrm{G.M}) = \frac{\sum_{i=1}^{k} f_i log(x_i)}{\sum_{i=1}^{k} f_i}$$

3. الوسط التوافقي: ويحسب بالصيغة:

$$\mathrm{H.M} = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

4. الوسيط: ويتم حسابه بالصيغة:

$$\tilde{X} = L_1 + \left[\frac{\frac{N}{2} - F_1}{f_{\tilde{X}}}\right] \times L_{\tilde{X}}$$

حيث

L_1 = الحد الأدنى لفترة الوسيط

F_1 = التكرار المتجمع الصاعد السابق لفترة الوسيط

$f_{\tilde{X}}$ = تكرار فترة الوسيط

$L_{\tilde{X}}$ = طول فترة الوسيط

N = مجموع التكرارات، $\sum_{i=1}^{k} f_i$، حيث k هو عدد الفترات[2]

[1] هذا المثال خاص بحساب الوسط الحسابي للبيانات ذات المشاهدات المتكررة، وسيتم حساب قيمة الوسط الحسابي، وباقي مقاييس النزعة المركزية، لبيانات جدول التوزيع التكراري في المثال القادم.

[2] لاحظ أن $k = c$ ، حيث أن c كان يرمز لعدد الفترات في الجدول التكراري في بداية تعريف جداول التوزيع التكراري.

وفترة الوسيط هي الفترة المناظرة لقيمة أول تكرار متجمع صاعد أكبر من أو يساوي رتبة الوسيط $\frac{\sum_{i=1}^{k} f_i}{2}$.

5. المنوال: وتوجد عدة صيغ لحساب المنوال في جداول التوزيع التكراري، أشهرها طريقة **الفروق لبيرسون** (Pearson's Differences)، والتي تعطى بالصيغة

$$\hat{X} = L_1 + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] \times L_{\hat{X}}$$

حيث

L_1 = الحد الأدنى لفترة المنوال

Δ_1 = قيمة الفرق بين تكرار فترة المنوال وتكرار الفترة التي تسبقها

Δ_2 = قيمة الفرق بين تكرار فترة المنوال وتكرار الفترة التي تليها

$L_{\hat{X}}$ = طول فترة المنوال

وفترة المنوال هي الفترة المناظرة لقيمة أكبر تكرار في الجدول التكراري.

6. التجزيئات: يمكن حساب قيم الربيعات، العشيرات، والمئينات، على الترتيب، من خلال الصيغ التالية:

$$\mathrm{Q}_i = L_1 + \left[\frac{(\frac{iN}{4}) - F_1}{f_{\mathrm{Q}}}\right] \times L_{\mathrm{Q}} \ \ , i = 1,2,3$$

و

$$\mathrm{D}_i = L_1 + \left[\frac{(\frac{iN}{10}) - F_1}{f_{\mathrm{D}}}\right] \times L_{\mathrm{D}} \ \ , i = 1,2, \dots, 9$$

و

$$\mathrm{P}_i = L_1 + \left[\frac{(\frac{iN}{100}) - F_1}{f_{\mathrm{P}}}\right] \times L_{\mathrm{P}} \ \ , i = 1,2, \dots, 99$$

حيث

L_1 = الحد الأدنى لفترة التجزيئ ، (الربيع، العشير، والمئين)

F_1 = قيمة التكرار المتجمع الصاعد السابق لفترة التجزيئ

f_{Q} ، f_{D} ، و f_{P} = قيمة تكرار فترة الربيع، العشير، والمئين على الترتيب

L_{Q} ، L_{D} ، و L_{P} = طول فترة الربيع، العشير، والمئين على الترتيب

ويتم تحديد فترة الربيع، العشير، والمئين من خلال حساب رتبها من المقادير $\frac{iN}{4}$ ، $\frac{iN}{10}$ ، و $\frac{iN}{100}$ على الترتيب أولا، ثم تحديد الفترة المناظرة لقيمة أول تكرار متجمع صاعد أكبر من رتبة الفترة، كما هو الحال في تحديد فترة الوسيط.

وسنقوم من خلال المثال القادم بحساب المقاييس الخاصة باستكشاف طبيعة النزعة المركزية لتوزيع البيانات، وذلك من خلال استخدام بيانات جدول التوزيع التكراري.

مثال (25.2): الجدول (16.2) هو جدول توزيع تكراري يمثل الدرجات النهائية لـ 70 طالبا من طلبة قسم الإحصاء في مقرر "الاستدلال الإحصائي I" في أحد الفصول الدراسية.

والمطلوب حساب المقاييس التالية:

1) الوسط الحسابي، 2) الوسط الهندسي، 3) الوسط التوافقي، 4) الوسط الفرضي، (اعتبر أن A = 60)، 5) المنوال، 6) الوسيط، 7) الربيع الأول، العشير الرابع، والمئين التسعون، 8) أوجد كلا من الوسيط والمنوال باستخدام التمثيل البياني لدرجات الطلبة، 9) وضح كيفية تمثيل الربيعات الثلاثة بيانيا باستخدام المضلع التكراري المتجمع النسبي.

جدول (16.2): جدول التوزيع التكراري لدرجات 70 طالبا في المثال (25.2).

الفترة المستمرة (الدرجة)	التكرار f (عدد الطلبة)
0 – 20	5
20 – 40	15
40 – 60	25
60 – 80	20
80 – 100	5

الحل:

لحساب المقاييس الثلاثة الأولى نحن بحاجة لإنشاء بعض الأعمدة الجديدة في جدول (16.2). هذه الأعمدة هي موضحة في جدول (17.2).

جدول (17.2): جدول التوزيع التكراري لدرجات الطلبة في المثال (25.2)، وحساب الحدود الخاصة بإيجاد بالأوساط.

1	2	3	4	5	6	7
الفترة المستمرة	**التكرار f**	**مركز الفترة x**	$f_i\, x_i$	$log\,(x_i)$	$f_i\, log\,(x_i)$	$f_i\,/x_i$
0 – 20	5	10	50	1	5	0.50
20 – 40	15	30	450	1.48	22.16	0.50
40 – 60	25	50	1250	1.70	42.47	0.50
60 – 80	20	70	1400	1.85	36.90	0.29
80 – 100	5	90	450	1.95	9.77	0.06
المجموع	70		3600		116.30	1.84

1) نستخدم الأعمدة 2، 3، و4 في جدول (17.2) لحساب الوسط الحسابي بالصيغة:

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = 51.43 \text{ درجة}$$

حيث أن عدد الفترات $k = 5$. ومعدل درجات الطلبة في هذا المقرر يدل على انخفاض مستوى التحصيل العلمي لهؤلاء الطلبة، إذ أن المستوى العام في المادة بالكاد يصل لدرجة النجاح، (وهي 50 درجة). وقد تدفعنا هذه النتيجة لإجراء المزيد من التحليل الإحصائي (الاستكشاف) للوقوف على الأسباب المؤدية لهذه النتيجة.

2) من الأعمدة 2، 3، 5، و6 نحصل على

$$\log(\text{G.M}) = \frac{\sum_{i=1}^{k} f_i \, log(x_i)}{\sum_{i=1}^{k} f_i} = \frac{116.30}{70} = 1.66$$

$$\text{G.M} = Anti\ (\log(\text{G.M})) = 10^{\text{G.M}} = 10^{(1.66)} = 45.87$$

أي أن الوسط الهندسي لدرجات الطلبة هو 45.87 درجة.

3) من الأعمدة 2، 3، و 7 يكون لدينا

$$\text{H.M} = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = \frac{70}{1.84} = 38.02$$

فيكون الوسط التوافقي لدرجات الطلبة هو 38.02 درجة.

ملاحظة: بالنظر لتوزيع البيانات في جدول (16.2) نرى بوضوح أن قيمة الوسط الحسابي هي الأقرب والأكثر منطقية من بين الأوساط الثلاثة إلى واقع درجات الطلبة. ولاحظ أيضا أن

$$\text{A.M} = \bar{X} = 51.43 \;>\; \text{G.M} = 45.87 \;>\; \text{H.M} = 38.02$$

4) الوسط الحسابي الفرضي للبيانات المبوبة يعطى بالصيغة

$$\bar{X}_g = \text{A} + \frac{\sum_{i=1}^{N} f_i d_i}{\sum_{i=1}^{k} f_i}$$

حيث

A = قيمة الوسط الفرضي

d_i = انحرافات قيم مراكز الفترات عن قيمة A، $i = 1, 2, \ldots, k$

وسنستخدم الجدول التالي، (الجدول (18.2))، وذلك لتبسيط حساب قيمة الوسط الفرضي، علما بأن قيمته هي $A = 60$:

$$\bar{X}_g = 60 + \frac{-600}{70} = 51.43$$

أي أن الوسط الحسابي الفرضي لدرجات الطلبة هو 51.43 درجة، وهو يساوي، في هذا المثال، الوسط الحسابي الاعتيادي تماما.

جدول (18.2): حساب القيم الخاصة بإيجاد الوسط الفرضي في المثال (25.2).

الفترة المستمرة	التكرار f	مركز الفترة x	$d_i = x_i - A$	$f_i\, d_i$
0 – 20	5	10	−50	−250
20 – 40	15	30	−30	−450
40 – 60	25	50	−10	−250
60 – 80	20	70	10	200
80 – 100	5	90	30	150
المجموع	70			−600

ملاحظات:

1. لا يشترط أن تساوي قيمة الوسط الحسابي الفرضي قيمة الوسط الحسابي دائما، إلا أن القيمتين تكونان متقاربتين عادة.

2. إذا ما قمنا بتغيير القيمة الافتراضية A فإن ذلك لا يغير من قيمة الوسط الفرضي للمتغير على الإطلاق.

5) من جدول (16.2)، نلاحظ أن فترة المنوال هي تلك الفترة المناظرة لأكبر تكرار (25)، وهي الفترة (60 – 40)، وبالتعويض في صيغة المنوال

$$\hat{X} = L_1 + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] \times L_{\hat{X}}$$

نحصل على

$$\Delta_1 = 25 - 15 = 10 \text{ , } \Delta_2 = 25 - 20 = 5$$

$$\hat{X} = 40 + \left[\frac{10}{10 + 5}\right] \times 20 = 53.33$$

وهذا يعني أن الدرجة 53.33 هي، تقريبا، الدرجة الأكثر تكرارا ضمن درجات الطلبة، ولاحظ أن هذه القيمة "تنزع" أو تقترب إلى وسط الدرجات.

6) لإيجاد الوسيط نحتاج لحساب التكرار المتجمع الصاعد كما هو موضح في جدول (19.2).

جدول (19.2): حساب التكرار المتجمع الصاعد لدرجات الطلبة في المثال (25.2).

الفترة المستمرة	التكرار f	التكرار المتجمع الصاعد	التكرار المتجمع الصاعد النسبي
0 – 20	5	5	% 7.14
20 – 40	15	20	% 28.57
40 – 60	25	45	% 64.29
60 – 80	20	65	% 92.86
80 – 100	5	70	% 100
المجموع	70		

نبدأ بحساب رتبة الوسيط $\frac{N}{2} = \frac{70}{2} = 35$ ، فتكون فترة الوسيط هي الفترة المناظرة لقيمة أول تكرار متجمع صاعد أكبر من 35، (وهو 45)، فتكون فترة الوسيط هي (60 – 40)، ويكون التكرار المتجمع الصاعد السابق لفترة الوسيط F_1 =20، وتكرار فترة الوسيط $f_{\tilde{X}} = 25$، وطول فترة الوسيط $L_{\tilde{X}} = 20$. وبالتعويض في صيغة الوسيط

$$\tilde{X} = 40 + \left[\frac{35 - 20}{25}\right] \times 20 = 52$$

أي أن وسيط درجات الطلبة في هذا المقرر هو 52 درجة، وهي قيمة قريبة جدا من قيمة الوسط الحسابي، مما يدل على عدم وجود قيمة متطرفة ضمن مفردات البيانات.

7) إن خطوات حساب التجزيئات هي مشابهة تماما لخطوات حساب الوسيط مع تغيير رتبة المقياس المطلوب. فلحساب قيمة الربيع الأول نقوم أولا بإيجاد فترة الربيع الأول من خلال حساب رتبة الربيع الأول، نضع $i = 1$

فنحصل على $17.5 = \frac{1\times70}{4} = \frac{iN}{4}$ ، وبالتالي تكون فترة الربيع الأول، (من جدول (19.2))، هي $(20-40)$ فيكون

$$\mathrm{Q}_1 = L_1 + \left[\frac{\left(\frac{N}{4}\right) - F_1}{f_{\mathrm{Q}}}\right] \times L_{\mathrm{Q}} = 20 + \left[\frac{17.5-5}{15}\right] \times 20 = 36.67$$

وهذا يعني أن ربع المفردات (الطلبة) قد تحصلوا على درجات أقل من 36 درجة تقريبا، وثلاثة أرباع الطلبة قد تحصلوا على درجات أعلى من 36 درجة.

نأتي لحساب قيمة العشير الرابع؛ رتبة العشير الرابع هي $28 = \frac{4\times70}{10} = \frac{iN}{10}$ ، وبالتالي لحساب فترة العشير الرابع، (من جدول (19.2))، هي $(40-60)$ ويكون

$$\mathrm{D}_4 = L_1 + \left[\frac{\left(\frac{4N}{10}\right) - F_1}{f_{\mathrm{D}}}\right] \times L_{\mathrm{D}} = 40 + \left[\frac{28-20}{25}\right] \times 20 = 46.4$$

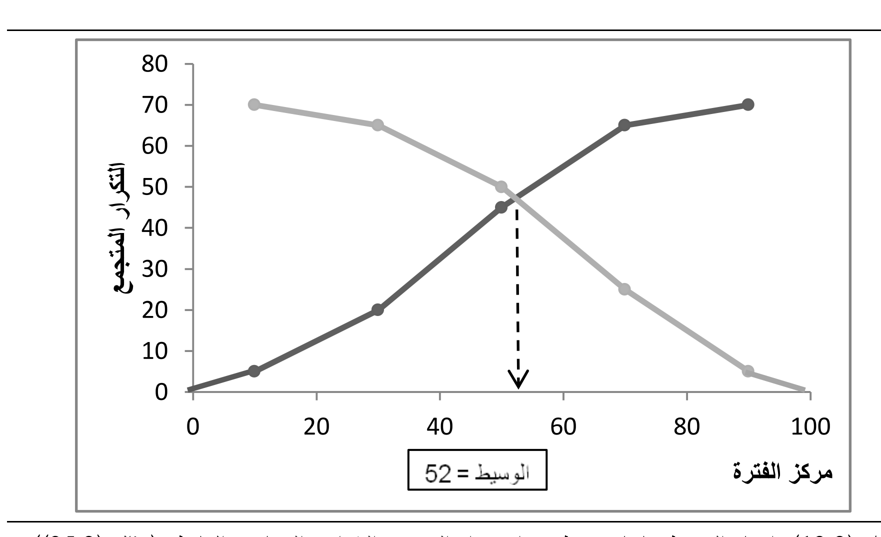
أي أن 40% من درجات الطلبة هو أقل من 46.4 درجة، بمعنى أن 40% من الطلبة هم دون درجة النجاح.

لحساب قيمة المئين التسعون نتبع نفس الخطوات السابقة، فنقوم بحساب رتبة المئين التسعون، وهي $= \frac{iN}{100}$ $63 = \frac{90\times70}{100}$ ، فتكون فترة المئين التسعون هي $(60-80)$، وقيمة المئين المطلوب هي

$$\mathrm{P}_{90} = L_1 + \left[\frac{\left(\frac{90N}{100}\right) - F_1}{f_{\mathrm{P}}}\right] \times L_{\mathrm{P}} = 60 + \left[\frac{63-45}{20}\right] \times 20 = 78$$

وهذا يعني أن 10% فقط من الطلبة تحصلوا على درجات أعلى من 78.

8) يمكننا إيجاد الوسيط عن طريق استخدام التمثيل البياني بأسلوبين، الأول هو عن طريق رسم كلا من المنحنى التكراري الصاعد والهابط، وتكون قيمة الوسيط عنده هي القيمة، (على المحور الأفقي X)، المقابلة لنقاط تقاطع المنحنيين، كما هو موضح في شكل (12.2).



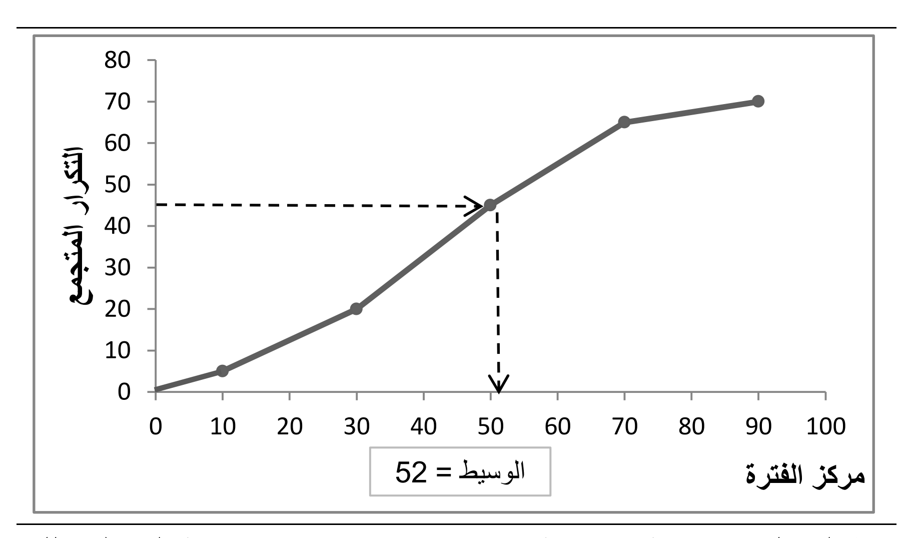
شكل (12.2): إيجاد الوسيط بيانيا عن طريق استخدام المنحنى التكراري الصاعد والهابط، (مثال (25.2)).

وأما الأسلوب الثاني فهو أن يتم رسم منحنى تكراري واحد فقط، (صاعد أو هابط)، وتكون قيمة الوسيط هي القيمة على محور X التي تلتقي مع قيمة التكرار المتجمع الصاعد المناظر لرتبة الوسيط، (وهي 45 في مثالنا)، على المحور العمودي Y عن طريق المنحنى التكراري، كما يتضح من شكل (13.2).

جدول (20.2): قيم التكرار المتجمع الصاعد والهابط للمثال (25.2).

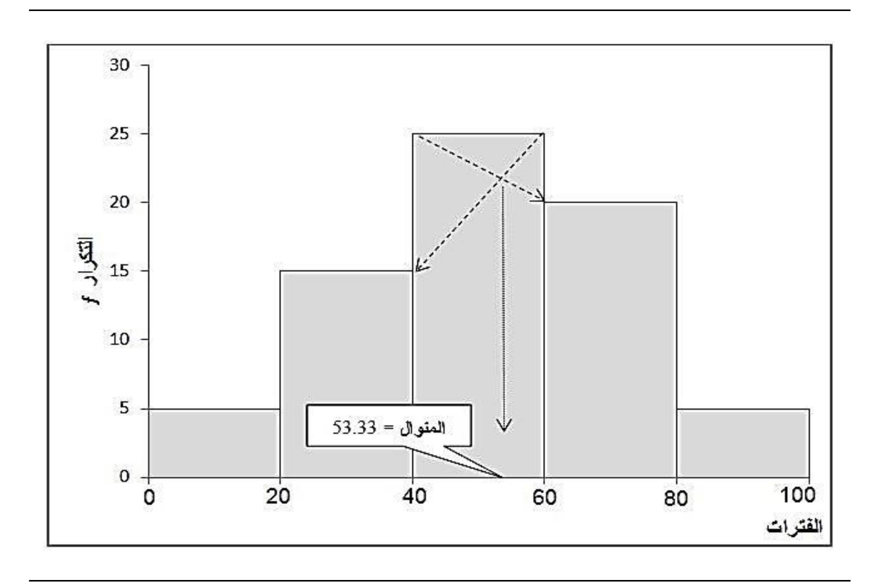
الفترة المستمرة	التكرار f	مركز الفترة x	التكرار المتجمع الصاعد	التكرار المتجمع الهابط
0 – 20	5	10	5	70
20 – 40	15	30	20	65
40 – 60	25	50	45	50
60 – 80	20	70	65	25
80 – 100	5	90	70	5
المجموع	70			

وجدول (20.2) يحتوي على القيم اللازمة لرسم المنحنى التكراري الصاعد والهابط.



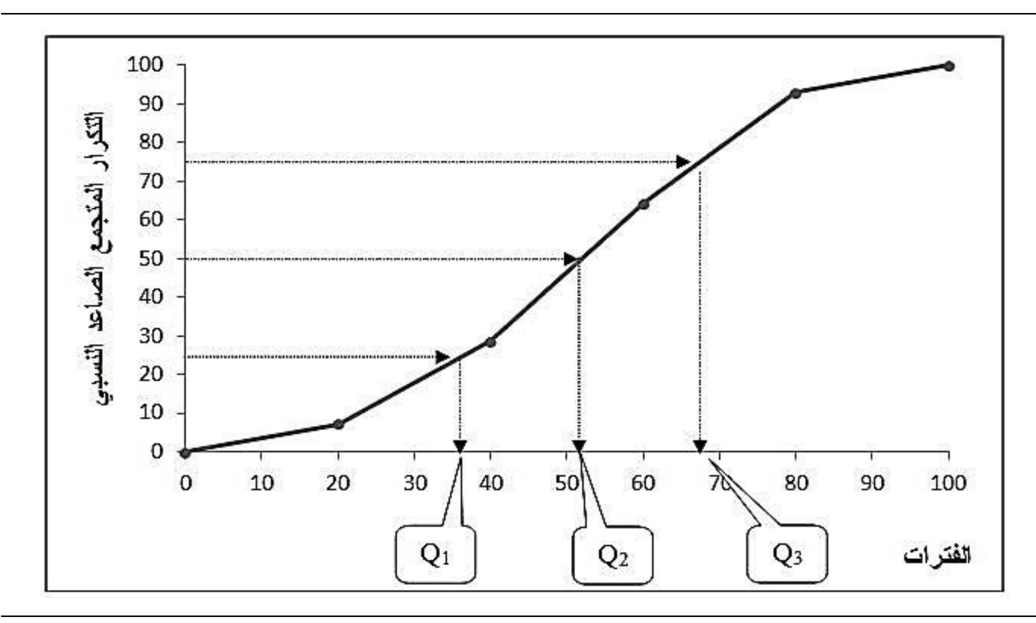
شكل (13.2): إيجاد الوسيط بيانيا عن طريق استخدام المنحنى التكراري الصاعد فقط، (مثال (25.2)).

أما المنوال فيتم حساب قيمته بيانيا عن طريق رسم المدرج التكراري لتوزيع البيانات، ثم اختيار فترة المنوال، وهي بديهيا الفترة المناظرة لأعلى عامود في المدرج. نقوم الآن برسم خط مستقيم يصل بين رأس زاوية العامود الأعلى اليمنى ورأس زاوية العامود الذي يسبقه (اليمنى أيضا)، ثم بنفس الكيفية ولكن عكسيا نرسم خط مستقيم يصل بين رأس زاوية العامود الأعلى اليسرى ورأس زاوية العامود الذي يليه (اليسرى). فتكون قيمة المنوال هي القيمة (على محور X) المناظرة لنقطة التقاء الخطان المستقيمان، كما هو موضح في شكل (14.2)، والذي تم استخدام جدول (16.2) في رسمه.



شكل (14.2): إيجاد قيمة المنوال بيانيا عن طريق استخدام المدرج التكراري، (مثال (25.2)).

9) لتمثيل الربيعات الثلاثة، نقوم أولا برسم المضلع القوسي النسبي كما هو موضح في شكل (15.2)، ثم نقوم بتحديد النسب الربيعية الثلاث 25%، 50%، و 75% على محور Y فتكون نقاط التقائها مع محور X ، (والذي يمثل الفترات)، عبر المضلع التكراري هي قيم Q_1، Q_2، وQ_3 على الترتيب. ولاحظ أننا استخدمنا قيم التكرار المتجمع الصاعد النسبي في جدول (19.2).



شكل (15.2): تحديد قيم الربيع الأول، الثاني، والثالث لدرجات الطلبة بيانيا على المضلع التكراري النسبي، (مثال (25.2)).

من شكل (15.2) يُلاحظ أن ربع الطلبة قد تحصلوا على درجات أقل من 36 درجة، ونصفهم تحصل على درجات أقل من 52 درجة (ويمكن القول أيضا أن نصف الطلبة تقريبا قد نجحوا على اعتبار أن 50 هي درجة النجاح)، وكذلك فإن ربع الطلبة تقريبا قد تحصلوا على درجات أعلى من 67 درجة. وبناءا على هذه المقاييس يمكن القول بأن أداء هؤلاء الطلبة في هذا المقرر لم يكن متميزا في العموم.

5.2 تمارين الفصل الثاني

تمرين (1.2): المشاهدات التالية تمثل قيم ضريبة الدخل المخصومة من موظفي إحدى الشركات الحكومية بالدينار الليبي:

185	170	130	175	180
248	195	201	145	160
111	127	210	177	230
165	167	174	220	155
161	129	145	150	225
185	190	171	224	180
245	120	195	179	153

والمطلوب تكوين جدول توزيع تكراري لهذه البيانات بطول فترة 25 مبتدأً بالقيمة 100 دينار.

تمرين (2.2): قامت إحدى القنوات التلفزيونية باستطلاع آراء مجموعة من المواطنين حول أداء قطاع الصحة في إحدى الدول العربية في السنوات الخمس الأخيرة فكانت آراؤهم كالتالي:

سيئ	مقبول	جيد	سيئ	مقبول	مقبول	مقبول	سيئ
مقبول	ممتاز	مقبول	جيد	جيد	ممتاز	جيد	مقبول
ممتاز	مقبول	سيئ	مقبول	مقبول	جيد	مقبول	جيد
مقبول							

والمطلوب:

أ. تنظيم الآراء السابقة في جدول توزيع تكراري، وكذلك حساب التكرار النسبي والنسبة المئوية للآراء بعد ترتيبها.

ب. ما نسبة المواطنين اللذين كانت إجاباتهم "جيد" أو "ممتاز" ؟.

تمرين (3.2): تم اختيار عينة مكونة من 50 مريضا من عدة مستشفيات يتناولون دواء لتنظيم ضربات القلب. وتم تسجيل الزمن اللازم لاستجابة كل مريض لهذا الدواء بالدقائق فكانت النتيجة كالتالي:

10	2.5	3	5	5.5	10	5.5	5.5	5	8.5
9.5	5	2.5	5.5	7	7	3.5	5	4.5	9
8	6.5	3	6	7	7.5	2	6	10	4.5
7.5	1.5	6.5	7.5	7	3.5	1.5	6.5	10	5
2.5	7	7.5	2	8.5	6.5	7	7	10.5	5

والمطلوب:

أ. كون جدول توزيع تكراري مستخدما 5 فترات منفصلة. (ملاحظة: يمكنك البدء بالفترة من1 إلى 2.9).

ب. أوجد التكرار المتجمع الصاعد، التكرار الصاعد النسبي، والنسبة التراكمية للبيانات.

تمرين (4.2): استخدم بيانات التمرين السابق (تمرين (3.2)) في:

أ. رسم مدرج تكراري. ب. رسم مضلع تكراري.

تمرين (5.2): الجدول التكراري التالي يمثل إجابات 100 مواطن من مدينة بنغازي طرح عليهم السؤال التالي: "إلى أي حد تعتقد أن ثورات ما يعرف بالربيع العربي قد ساهمت في تحرر الشعوب العربية؟".

إجابة المواطن	التكرار f
كثيرا	34
إلى حد ما	18
ليس كثيرا	19
لا على الإطلاق	14
لست متأكد	15

والمطلوب استخدام كلا من الأعمدة البيانية والقطاعات الدائرية لتمثيل الإجابات.

تمرين (6.2): الجدول التالي يوضح توزيع أعداد حوادث المرور حسب نوع المخالفة (الأعمدة) وحسب تصنيف المصابين ما إذا توفوا أم لا (الصفوف) في إحدى المدن لسنة 2013:

	نوع المخالفة					
تصنيف المصاب ↓	**اجتياز إشارة حمراء**	**سرعة عالية**	**خلل في المركبة**	**تحت تأثير مخدر**	**بدون رخصة قيادة**	**سير عكس الاتجاه**
توفى	121	156	89	32	47	25
لم يتوفى	213	187	109	44	36	15

والمطلوب تمثيل توزيع أعداد هذه الحوادث بيانيا كالتالي:

أ. حسب نوع المخالفة وتصنيف المصاب باستخدام الأعمدة البيانية.

ب. حسب نوع المخالفة فقط باستخدام القطاعات الدائرية.

ج. حسب تصنيف المصاب باستخدام القطاعات الدائرية.

تمرين (7.2): تم رصد معدلات درجات الحرارة المئوية عند قمة أحد الجبال في شمال أوروبا خلال بداية فصل الشتاء لمدة 40 يوما فكانت النتيجة كما هو موضح أدناه:

0	−4	1	−1	−7.5	−6	6	−9	0	2.5
−4	−9	−1	−2	−6.5	0	−8	3.5	−2	4.5
5	8	4	−5	0	−1	7	−4	−1	7
−1	−1	−3	2	−1	−2	−7	3	3	1

والمطلوب حساب:

أ. الوسط الحسابي، ب. الوسيط، و ج. المنوال لهذه البيانات.

تمرين (8.2): الجدول التكراري التالي يلخص توزيع أعداد 85 شخصا (أعمارهم 30 سنة فأقل) يعانون من مرض السكر وتم توزيعهم حسب الفئة العمرية.

التكرار f	الفترة (الفئة العمرية)
7	0 − 5
18	5 − 10
26	10 − 15
19	15 − 20
10	20 − 25
5	25 − 30

والمطلوب حساب:

أ. الوسط الحسابي.

ب. الوسط الهندسي.

ج. الوسط التوافقي.

د. الوسط الفرضي (اعتبر A=15).

هـ. المنوال.

و. الوسيط.

ز. الربيع الثالث.

ح. العشير الثالث.

ط. المئين الخامس والسبعون.

الفصل الثالث

التحليل الاستكشافي للبيانات – الجزء الثاني
(Exploratory Data Analysis (EDA) – Part Two)

1.3 مقاييس التشتت (Measurements of Dispersion)

1.1.3 المدى (Range)

2.1.3 المدى الربيعي (Interquartile Range, IQR)

3.1.3 المدى المئيني (10 – 90) (Percentile Range)

4.1.3 الانحراف المتوسط (Mean Deviation)

5.1.3 الانحراف المعياري (Standard Deviation)

6.1.3 التباين (The Variance)

7.1.3 معامل الاختلاف أو التشتت (Coefficient of Variation or Dispersion)

8.1.3 مقاييس التشتت لبيانات الجداول التكرارية
(Measurements of Dispersion for Tabulated Data)

9.1.3 الدرجات المعيارية (Standard Units or Z-scores)

2.3 العزوم، الالتواء، والتفرطح (Moments, Skewness, and Kurtosis)

1.2.3 العزوم (Moments)

2.2.3 الالتواء والتفرطح (Skewness and Kurtosis)

3.3 بعض الرسومات البيانية الإضافية (Some Additional Graphical Tools)

1.3.3 شكل الصندوق (The Boxplot)

2.3.3 شكل الساق والورقة (The Stem-leaf plot)

4.3 تمارين الفصل الثالث

1.3 مقاييس التشتت (Measurements of Dispersion)

في الفصل السابق تم مناقشة أهمية مقاييس النزعة المركزية في فهم طبيعة توزيع وتمركز البيانات، وسنتناول في هذا الفصل أهمية مقاييس التشتت في وصف انتشار قيم هذه البيانات حول مركزها، وكيفية وصف التغير في القيم بالنظر إلى نزعتها المركزية.

فالتغير عادة ما يكون ملاحظا في أي مجموعة من البيانات بغض النظر عن الوحدات المستخدمة في القياس أو طبيعة مفردات الظاهرة، فكل ما يحيط بنا وكل ما نتعامل معه في الحياة اليومية وكل ما نرصده أو نراقبه يتغير، فالأعمار والأوزان والأطوال والأسعار والإنتاج وغيرها، كلها تتغير من زمن لآخر ومن مكان لآخر ومن مشاهدة لأخرى، (ولهذا نسميها متغيرات). ولأنه ببساطة لن يكون من المنطقي أن تأخذ كل مفردة من البيانات نفس القيمة ضمن المتغير الواحد، لأنه لن يسمى في هذه الحالة متغيرا بل ثابتا[1]. وعمليا، فإن عدم وجود التغير في البيانات يقلل من الحاجة لاستخدام الإحصاء سواء في الاستكشاف أو التحليل الاستدلالي.

من ناحية أخرى فقد نصادف مجموعتين، أو أكثر، من المتغيرات التي يكون لها نفس قيمة الوسط أو الوسيط، ولكن مفرداتها تنتشر بمستويات مختلفة، فمثلا إذا ما تم حساب معدل درجات الحرارة في مدينتين متجاورتين في فترة زمنية معينة في فصل الصيف وكان متساويا، فإن هذا لا يعني بالضرورة أن درجات الحرارة اليومية في كلا المدينتين كانت متساوية، ولهذا تكون الحاجة هنا أكبر لاستخدام مقاييس التشتت. وعمليا فإن أي دراسة وصفية أو استكشافية لمجموعة من البيانات لا يجب أن تخلوا من كلا النوعين من المقاييس؛ النزعة المركزية والتشتت.

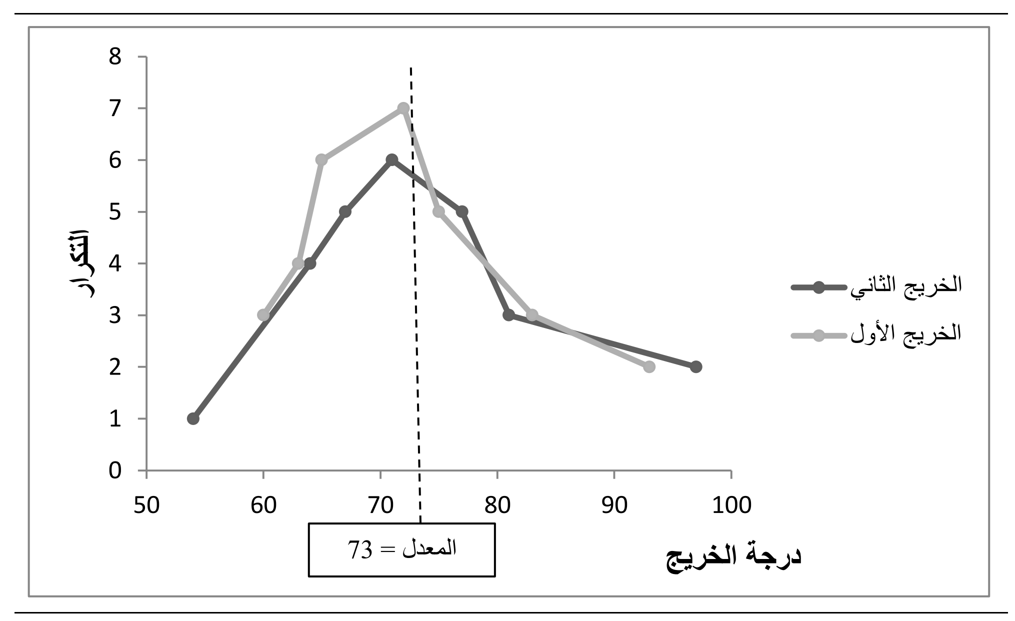
إن التشتت ببساطة هو الدرجة التي تميل إليها مفردات البيانات إلى الانتشار أو التوزع حول قيمة وسطية (مركزية)، ولتوضيح هذا المفهوم بصورة عملية وتسليط الضوء على أهمية مقاييس التشتت في استكشاف البيانات لنأخذ المثال التوضيحي التالي:

لنفرض أن لجنة علمية في إحدى الجامعات تريد المفاضلة بين خريجَين من المتقدمين لبرنامج الدراسات العليا في إحدى التخصصات على اعتبار أنهما قد اجتازا المقابلة الشخصية وتحصلا على أفضل النتائج الدراسية ضمن باقي المتقدمين. من البديهي أن يتم اختيار الخريج الذي تحصل على أعلى معدل (متوسط) دراسي، ولكن كيف يكون الحال إذا كان الخريجان قد تحصلا على نفس المعدل الدراسي العام؟، ولمزيد من التوضيح لنعتبر الشكل التالي، (شكل (1.3))، والذي يوضح توزيع درجات الخريجَين على مر عدة فصول دراسية قبل التخرج.

نلاحظ من توزيع الدرجات في الشكل المرفق أن كلا الخريجَين لهما نفس المعدل إلا أنهما يختلفان في توزع أو انتشار (تشتت) درجاتهما خلال الفصول الدراسية، وهذا يعني أن الحكم بقبول أحدهما بناءا على مقياس أعلى معدل، (كمقياس للنزعة المركزية)، لا يكفي وحده. وهنا يبرز سؤال آخر؛ هل تختار اللجنة الخريج الذي تحصل على أعلى معدل، ضمن الفصول الدراسية ككل، وهو الخريج الثاني (97 درجة)؟ والذي تحصل في نفس الوقت

[1] في بعض الحالات الخاصة قد ينتج عن التجربة أو الدراسة مجموعة ثابتة من القيم، كأن يتم رصد كميات إنتاج مصنع لعدد من الأيام علما بأن إنتاجه ثابت في اليوم الواحد، عندئذ لن يكون هنالك داعي لاستخدام الأساليب الإحصائية الاعتيادية من مقاييس نزعة مركزية وتشتت لمراقبة تغير المفردات.

على أدنى درجة (54 درجة)؟، أم تختار الخريج الأول والذي لم يصل إلى أعلى درجة إلا أنه لم يصل أيضا إلى أدنى درجة؟



شكل (1.3): شكل انتشار يوضح توزيع درجات الخريجَين على مر عدة فصول دراسية.

إن الإجابة على هذه التساؤلات يتضمن استخدام مقاييس تهتم بدراسة انتشار مفردات هذه البيانات، وهي ما يعرف بمقاييس التشتت. أما حاليا، بالنسبة لمثالنا التوضيحي فيمكننا القول أن درجات الخريج الأول هي أقل تشتتا من درجات الخريج الثاني، وهذا يعني أن درجاته أقرب للوسط (المركز)، وعموما فإن الدرجات الأقل تشتتا، (داخل منطقة الدرجات العالية)، تدل على أن مستوى الخريج في العموم أفضل.

إضافة لما سبق، فإن دراسة الوسط أو الأوساط لمجموعة من البيانات قد لا يكون أحيانا ذو أهمية، أو على الأقل نقول أنه يفقد أهميته إذا ما كان تشتت هذه البيانات عن قيمة الوسط كبير جدا[1].

ومن الناحية التطبيقية، فإن بعض مقاييس التشتت يتعامل مع وحدات قياس المفردات مباشرة أو يتعامل مع قيم المفردات بشكل مباشر وهي مقاييس مطلقة، والبعض الآخر يتعامل مع نسب بين المقاييس المحسوبة نفسها وهي مقاييس نسبية. وفيما يلي سنقوم بدراسة أهم مقاييس التشتت التي تستخدم ضمن إطار الإحصاء الاستكشافي أو الوصفي. وكما هو الحال في الفصل السابق، سنبدأ مع البيانات المفردة (غير المبوبة).

1.1.3 المدى (Range)

لاحظنا في الجزء (4.2) في الفصل الثاني، كيفية استخدام المدى مع البيانات المفردة بغرض تكوين جدول التوزيع التكراري، والقانون هو نفسه هنا، فالمدى يعد من أبسط المقاييس التي تعنى بقياس التغير ضمن مفردات البيانات.

[1] وهذا قد يبنى عليه اعتبارات أخرى خاصة بعملية المعاينة وتتعلق بشكل كبير بالاستدلال الإحصائي.

تعريف (1.3): **المدى:** المدى (R) لمفردات البيانات هو الفرق بين أكبر قيمة في هذه المفردات (x_{max}) وأصغر قيمة (x_{min})، أي

$$R = x_{max} - x_{min}$$

والبساطة في استخدام المدى كمقياس تكمن في سهولة تفسير قيمته من خلال ما تمثله مفردات المتغير. فمثلا، حساب المدى لأجور مجموعة من العمال في مصنع ما يوضح مدى الاختلاف بحيث أن القيمة الكبيرة للمدة تعني عدم وجود تجانس بين ما يتقاضاه هؤلاء العمال في هذا المصنع، وكذلك فإن حساب المدى لدرجات مجموعة من الطلبة المختلفين في التخصص في مقرر عام يعكس درجة التفاوت أو الاختلاف في "الاهتمام" بهذا المقرر بين هؤلاء الطلبة.

وعلى الرغم من أن المدى كمقياس يحدد الانتشار الكلي لمفردات البيانات، إلا أن حسابه يعتمد على مفردتين فقط وهما المفردتين الأكثر تطرفا (بعدا) ضمن قيم المفردات، وهذا في الحقيقة يعد من أبرز عيوب المدى كمقياس للتشتت. إذ أنه بزيادة قيمة x_{min} أو x_{max} أو كلاهما مع "استقرار" باقي القيم فإن قيمة المدى سوف تتغير بشكل ملحوظ، ولن تكون بحد ذاتها ذات فائدة كبيرة في وصف انتشار البيانات.

مثال (1.3): أوجد المدى للمفردات التالية والتي تمثل أطوال عشرة أطفال (بالمتر) في المرحلة الابتدائية.

1.15	0.90	0.85	1.21	0.93	0.74	0.80	0.98	1.25	1.10

الحل:

$$R = x_{max} - x_{min} = 1.25 - 0.74 = 0.51 \text{ متر}$$

وهذا يعني وجود تفاوت كبير بين أطوال هؤلاء الأطفال يصل إلى نصف متر، مما يدل على اختلاف أعمارهم أو وجود عوامل أخرى، (عامل وراثي مثلا).

2.1.3 المدى الربيعي ((Interquartile Range,(IQR))

هذا المقياس يعتمد على حساب المدى، إلا إنه لا يستخدم القيم الصغرى والكبرى لمفردات البيانات بل يستخدم الربيع الأول والربيع الثالث عوضا عنها، ويسمى أحيانا **بالانحراف الربيعي** (Quartile Deviation).

تعريف (2.3): **المدى الربيعي:** هو الفرق بين قيمتي الربيع الأول والثالث لمفردات البيانات:

$$\mathrm{IQR} = \mathrm{Q}_3 - \mathrm{Q}_1$$

وإذا ما قمنا بقسمة المدى الربيعي على 2 فإننا نحصل على مقياس آخر هو **نصف المدى الربيعي** (Semi-Interquartile Range,(SIQR))، والذي يعد أكثر استخداما من المدى الربيعي[1].

تعريف (3.3): **نصف المدى الربيعي:** هو نصف الفرق بين الربيع الأول والثالث لمفردات البيانات:

$$\mathrm{SIQR} = \frac{\mathrm{Q}_3 - \mathrm{Q}_1}{2}$$

[1] لأن قسمة المدى الربيعي على 2 تجعله أقرب إلى مركز البيانات.

ونلاحظ أن المقياسان الأخيران يتغلبان بصورة كبيرة على العيب الأساسي في مقياس المدى وهو شدة التأثر بالقيم المتطرفة، لأنهما يعتمدان على قيم تبتعد عن الأطراف.

مثال (2.3): أوجد نصف المدى الربيعي للبيانات التالية:

1	2	3	6	9	13	14	16	21	25	27	28	100

الحل:

حيث أن عدد المفردات هو فردي فإن رتبة الربيع الأول هي $\frac{N+1}{4} = \frac{14}{4} = 3.5 \cong 4$ أي المفردة الرابعة فيكون $Q_1 = 6$، كما هو موضح أدناه:

<table>
<tr><td colspan="13">Q_2 (14)</td></tr>
<tr><td colspan="3">25% من المفردات</td><td>Q_1</td><td colspan="9">75% من المفردات</td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>6</td><td>9</td><td>13</td><td>14</td><td>16</td><td>21</td><td>25</td><td>27</td><td>28</td><td>100</td></tr>
<tr><td colspan="9">75% من المفردات</td><td>Q_3</td><td colspan="3">25% من المفردات</td></tr>
</table>

ورتبة الربيع الثالث هي $\frac{3N+1}{4} = \frac{40}{4} = 10$ أي المفردة العاشرة فيكون $Q_3 = 25$. ثم نقوم الآن بحساب نصف المدى الربيعي

$$SIQR = \frac{Q_3 - Q_1}{2} = \frac{25 - 6}{2} = 9.5$$

وإذا ما اعتبرنا أن

$$\frac{Q_1 + Q_3}{2} = \frac{6 + 25}{2} = 15.5$$

هي قيمة وسطية (مركزية) للبيانات، فيمكننا القول أن 50% من قيم البيانات (وهي القيم من $Q_1= 6$ إلى $Q_3= 25$) تنتشر في النطاق أو الفترة 15.5 ± 9.5 ، ويمكن ملاحظة أن مقياس نصف المدى الربيعي لم يتأثر بالقيمة المتطرفة (100) كما كان سيكون الحال مع مقياس المدى ($R = 100 - 1 = 99$).

3.1.3 المدى المئيني (10 – 90) (Percentile Range)

وهو يشبه المدى الربيعي من حيث طريقة التطبيق، ويفضل استخدامه عند التعامل مع قواعد البيانات الكبيرة.

تعريف (4.3): **المدى المئيني (10 – 90)**: هو الفرق بين قيمتي المئين العاشر والمئين التسعون:

$$R_{P_{90}-P_{10}} = P_{90} - P_{10}$$

4.1.3 الانحراف المتوسط (Mean Deviation)

لاحظنا في المقاييس السابقة أنها تستخدم أو تعتمد في حسابها على بعض القيم فقط ولا تأخذ في الاعتبار كل قيم المفردات، وهذا قد يسبب نقصا في المعلومات التي يمكن الاستفادة منها حول طبيعة انتشار البيانات. لذلك

نكون بحاجة لاستخدام مقاييس تشمل في طريقة حسابها على كل قيم البيانات. والانحراف المتوسط هو أحد هذه المقاييس.

تعريف (5.3): **الانحراف المتوسط:** الانحراف المتوسط للمفردات $x_1, x_2, \ldots, x_N$ يعرف بأنه متوسط الانحرافات المطلقة عن الوسط الحسابي لهذه المفردات ويرمز له بالرمز M.D، ويعطى بالصيغة:

$$\mathrm{M.D} = \frac{\sum_{i=1}^{N}|x_i - \bar{X}|}{N}$$

وقد يسمى الانحراف المتوسط **بالانحراف المتوسط المطلق** (Absolute Mean Deviation) أحيانا.

مثال (3.3): أوجد الانحراف المتوسط لهذه المفردات 8، 6، 2، 11، 3 .

الحل:

الوسط الحسابي لهذه المفردات هو $\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{30}{5} = 6$ ، فيكون الانحراف المتوسط

$$\mathrm{M.D} = \frac{|3-6| + |11-6| + |2-6| + |6-6| + |8-6|}{5} = \frac{14}{5} = 2.8$$

وهذا يعني أن انتشار المفردات حول وسطها الحسابي، بغض النظر عن كون المفردة أكبر من الوسط أو أقل منها (على اعتبار أنه تم استخدام القيمة المطلقة)، هو بمعدل $3 \cong 2.8$ وحدات تقريبا.

ملاحظة: في بعض الحالات قد تستخدم صيغة عامة للانحراف المتوسط بحيث يتم قياس انحراف المفردات عن أي قيمة مركزية a . بمعنى أن:

$$\mathrm{M.D} = \frac{\sum_{i=1}^{N}|x_i - a|}{N}$$

علما بأنه عندما تكون $a = \tilde{X}$ فإن الانحراف المتوسط يأخذ أقل قيمة له فيما إذا كانت a هي قيمة وسطية أخرى.

5.1.3 الانحراف المعياري (Standard Deviation)

يعد مقياس الانحراف المعياري من أشهر وأهم مقاييس التشتت، ويشبه الانحراف المتوسط من حيث أنه أيضا يقيس انتشار المفردات حول وسطها الحسابي، إلا أنه لا يهمل الانحرافات السالبة، (كما هو الحال مع الانحراف المتوسط)، بل يقوم بتربيع هذه الانحرافات، وهذا الإجراء يكون أقرب إلى الدقة من الناحية الجبرية والمنطقية.

تعريف (6.3): **الانحراف المعياري:** الانحراف المعياري للمفردات $x_1, x_2, \ldots, x_N$ ، والذي يرمز له بالرمز S.D، يعرف بأنه الجذر التربيعي لمتوسط مربعات انحرافات قيم المفردات عن وسطها الحسابي، ويعطى بالصيغة:

$$\mathrm{S.D} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{X})^2}{N}}$$

ملاحظة: الكمية تحت الجذر في صيغة الانحراف المعياري، (تعريف (6.3))، يمكن معالجتها لتصبح بالصورة:

$$\frac{\sum_{i=1}^{N}(x_i-\bar{X})^2}{N}=\frac{\sum_{i=1}^{N}(x_i^2+\bar{X}^2-2\bar{X}x_i)}{N}$$

$$=\frac{\sum_{i=1}^{N}x_i^2}{N}+\frac{N\bar{X}^2}{N}-2\bar{X}\frac{\sum_{i=1}^{N}x_i}{N}$$

$$=\frac{\sum_{i=1}^{N}x_i^2}{N}+\bar{X}^2-2\bar{X}^2$$

$$=\frac{\sum_{i=1}^{N}x_i^2}{N}-\bar{X}^2$$

$$=\frac{\sum_{i=1}^{N}x_i^2}{N}-\left(\frac{\sum_{i=1}^{N}x_i}{N}\right)^2$$

وبالتالي يصبح

$$\text{S.D}=\sqrt{\frac{\sum_{i=1}^{N}x_i^2}{N}-\left(\frac{\sum_{i=1}^{N}x_i}{N}\right)^2}$$

وهذه الصيغة عادة ما تستخدم بصورة تطبيقية لسهولة التعامل مع مكوناتها.

مثال (4.3): أوجد الانحراف المعياري للبيانات التالية (جدول (2.3)) والتي تمثل أطوال خمسة أنواع من الحشرات (بالسنتيمتر) في سن التزاوج؛ 6، 4، 3، 2، 1 .

الحل:

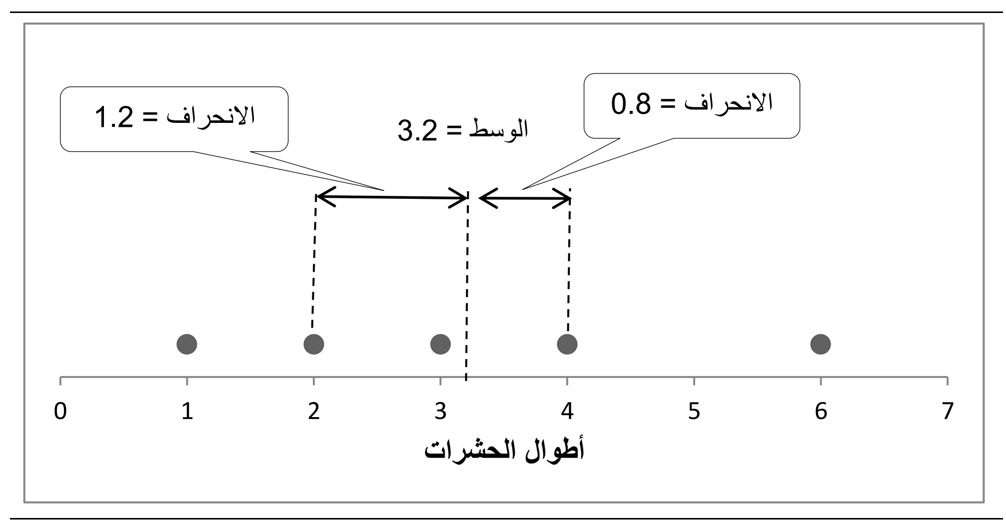
لتبسيط الحسابات نقوم باستخدام الجدول التالي:

جدول (2.3): حساب الكميات اللازمة لإيجاد الانحراف المعياري، (مثال (4.3)).

x	1	2	3	4	6	المجموع	16
x^2	1	4	9	16	36		66

$$\text{S.D}=\sqrt{\frac{\sum_{i=1}^{N}x_i^2}{N}-\left(\frac{\sum_{i=1}^{N}x_i}{N}\right)^2}=\sqrt{\frac{66}{5}-\left(\frac{16}{5}\right)^2}=\sqrt{2.96}=1.72\ \text{سم}$$

ولتوضيح "مفهوم" هذه النتيجة سنستخدم الرسم النقطي التالي، (شكل (2.3)). نلاحظ من الشكل (2.3) أن انحراف المفردة $x = 2$ عن الوسط الحسابي هو (1.2) وانحراف المفردة $x = 4$ عن الوسط الحسابي هو (0.8)، ويمكن حساب انحرافات باقي قيم المفردات عن الوسط الحسابي بالمثل أيضا. فالانحراف المعياري يقيس "متوسط" ابتعاد كل المفردات عن الوسط الحسابي حيث أننا لن نستطيع ملاحظة ابتعاد كل نقطة بمفردها عن المركز (الوسط الحسابي) بشكل رقمي، ولهذا نكون بحاجة لمقياس واحد، (وهو الانحراف المعياري)، لتلخيص وقياس هذا التشتت عن المركز. وهكذا فإننا نستطيع القول بأن الاختلاف في أطوال هذه الحشرات هو بمقدار 1.72 سم عن معدل الطول 3.2 سم.



شكل (2.3): الرسم النقطي لأطوال الحشرات في مثال (4.3).

ملاحظات (حول الانحراف المعياري):

1. لابد من فهم ما تمثله البيانات بصورة صحيحة لكي يتسنى لنا الاستفادة من المعلومة المكتسبة عن طريق حساب الانحراف المعياري بشكل عملي. وكذلك فإن التأكد من وحدة القياس، (متر، كيلوجرام، وات، يوم، درجة،...)، المستخدمة في البيانات هو أمر مهم جدا، فمثلا إذا كانت قيمة الانحراف المعياري هي 2 سنة فهذا مكافئ لقيمة الانحراف المعياري 24 شهر.

2. لاحظ أنه بزيادة قيم المفردات في الأطراف، (وجود قيم متطرفة)، فإن قيمة الانحراف المعياري ستزيد بدورها والعكس صحيح، أي أنه كلما قلت قيمة الانحراف المعياري كان هذا معناه أن قيم المفردات قريبة من الوسط أكثر (في المعدل).

3. في بعض الحالات قد يكون الحصول على قيمة صغيرة للانحراف المعياري هو المؤشر الأهم، أو المعلومة المطلوب التحقق منها، فمثلا عند مراقبة إنتاج مصنع ما فإن الحصول على قيمة صغيرة للانحراف المعياري لمواصفات السلع المنتجة، (من حيث الوزن أو الحجم أو المواصفات القياسية أو ...)، يعني أن أخطاء التصنيع هي تحت السيطرة.

4. يستخدم الوسط الحسابي $\bar{X}$ عادة كمركز لحساب الانحراف المعياري بدلا من أي قيمة وسطية أخرى لأن استخدام $\bar{X}$ يعطي أقل قيمة للانحراف المعياري فيما لو استخدمت أي قيمة أخرى[1].

5. عندما تكون قيمة الانحراف المعياري مساوية للصفر، $(S.D = 0)$، فهذا يعني عدم وجود أي تشتت أو انحراف لكل القيم عن الوسط الحسابي، وهذا يعني أنه تم حساب الانحراف المعياري لمجموعة من القيم الثابتة، أو حسابه لمفردة واحدة[2]. ولهذا فإنه يكون $S.D \geq 0$ دائما.

---

[1] يمكنك التأكد من ذلك رقميا باستخدام أي مجموعة بسيطة من البيانات وحساب الانحراف المعياري باستخدام قيمة الوسيط أو أي وسط آخر عوضا عن الوسط الحسابي في الصيغة الأصلية.

[2] إذا كان a ثابت، فإن؛ $S.D(a) = \sqrt{\frac{\sum_{i=1}^{1}(a-a)^2}{1}} = 0$ ، لأن الوسط الحسابي للثابت (المفردة الواحدة) هو نفسه.

6. الصيغة الموضحة في التعريف (6.3) تستخدم فقط عند التعامل مع البيانات التي تمثل المجتمع، وهذا أمر نادر عمليا، إذ أن أغلب حسابات المقاييس الإحصائية تكون لعينات، (حجمها n مثلا)، مسحوبة من المجتمع. في هذه الحالة تكون صيغة الانحراف المعياري معرفة بالصورة

$$\mathrm{S.D}=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n-1}}$$

والصيغة الحسابية المستخدمة عمليا هي

$$\mathrm{S.D}=\sqrt{\frac{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}{n(n-1)}}$$

7. يستخدم الانحراف المعياري أيضا للمقارنة بين درجات تشتت مجموعتين أو أكثر من البيانات أو متغيرات ضمن البيانات[1]، ويفضل أن يكون لها متوسطات قريبة من بعضها البعض[2]. والمثال القادم يوضح الصورة.

مثال (5.3): البيانات في جدول (3.3) تمثل أوزان مجموعة من أرغفة الخبز (بالجرام) تم جمعها من مخبزين مختلفين A و B. والمطلوب استخدام مقياس الانحراف المعياري لتوضيح أي من المخبزين أكثر التزاما بالمواصفات القياسية لوزن الخبز التي وضعتها الدولة، (وهو 100 جرام للرغيف الواحد)، حسب التسعيرة المقررة.

جدول (3.3): أوزان أرغفة الخبز الخاصة بالمخبزين A و B في المثال (5.3).

x_A	90	95	98	102	101	97	98	99	96
x_B	95	90	91	95	87	100	109	99	102

الحل:

لدينا بالنسبة للمخبز A:

$$\sum_{i=1}^{9}x_{\mathrm{A}i}=876\ ,\sum_{i=1}^{9}x_{\mathrm{A}i}^2=85364\ ,\bar{X}_{\mathrm{A}}=97.33\ ,\mathrm{S.D_A}=3.33$$

وبالنسبة للمخبز B:

$$\sum_{i=1}^{9}x_{\mathrm{B}i}=868\ ,\sum_{i=1}^{9}x_{\mathrm{B}i}^2=84086\ ,\bar{X}_{\mathrm{B}}=96.44\ ,\mathrm{S.D_B}=6.43$$

وهكذا نلاحظ أنه على الرغم من أن كلا المخبزين ينتج أرغفة خبز بمعدل وزن متقارب جدا وقريب نوعا ما من الوزن القياسي، إلا أن المخبز B إنتاجه للخبز (أوزان الخبز فيه) يشوبها تشتت أو اختلاف أكبر ($\mathrm{S.D_A} < \mathrm{S.D_B}$)، وهذا يعني أنه أقل التزاما أو دقة في الإنتاج، لأن البعض قد يحصل على أرغفة خبز ناقصة الوزن بشكل ملحوظ.

[1] على أن تكون لها نفس وحدات القياس.

[2] لأن ذلك يزيد من دقة المقارنة، اعتبارا أن مقارنات انحرافات القيم ستكون عن مراكز قريبة من بعضها البعض.

8. باستخدام كلا من الوسط الحسابي والانحراف المعياري يمكن الحصول على وصف جيد لتوزيع مفردات البيانات من خلال النظرية الشهيرة التالية:

نظرية (1.3): **نظرية تشيبيتشف**[1] (Chebyshev's Theorem):

يجب أن يقع على الأقل الجزء $\left(1-\frac{1}{k^2}\right)$ من مفردات البيانات في الفترة $\bar{X} \pm k(\mathrm{S.D})$ ، حيث k هي أي قيمة ثابتة.

ولتوضيح النظرية (1.3) نأخذ القيمة $k = 2$ مثلا فنحصل على $1-\frac{1}{2^2}=\frac{3}{4}$ ، عندئذ يقال أن $\frac{3}{4}$ أو 75% من المفردات على الأقل لابد أن يقع في الفترة $\bar{X} \pm 2(\mathrm{S.D})$ ، أي لابد أن يقع على بعد 2(S.D) على جانبي الوسط الحسابي.

وكذلك إذا ما وضعنا $k = 3$ فإن $1-\frac{1}{3^2}=\frac{8}{9}$ ونقول أن $\frac{8}{9}$ أو 88.9% من المفردات على الأقل لابد أن يقع في الفترة $\bar{X} \pm 3(\mathrm{S.D})$.

ويجب التنويه هنا إلى أن نظرية تشيبيتشف لا تكون ذات فائدة عندما توضع $k = 1$ لأنه عندئذ يكون $1-\frac{1}{1^2}=0$ ، وهذا يعني أنه لن توجد مفردات في الفترة $\bar{X} \pm (\mathrm{S.D})$ وهذا غير منطقي. وإضافة إلى ذلك، فإن النظرية لا تحدد لنا مقدار المفردات الذي يقع في فترة معينة بالضبط، بل تحدد "على الأقل" الجزء الذي سيقع في تلك الفترة.

مثال (6.3): في إحدى المدن الساحلية تم تسجيل نسبة الرطوبة لمدة 1080 يوما خلال فصل الصيف على مر عدة سنوات، فكان متوسط نسبة الرطوبة هو 120 درجة بانحراف معياري 8 درجات. استخدم نظرية تشيبيتشف لتحديد الفترة التي تحتوي على الأقل 810 أيام.

الحل:

لدينا $\frac{3}{4}=\frac{810}{1080}=1-\frac{1}{k^2}$ ، وهذا يعني أن $k = 2$ ، وبالتالي فإن الفترة التي تحتوي على الأقل $\frac{3}{4}$ من المفردات (الأيام) هي $\bar{X} \pm 2(\mathrm{S.D}) = 120 \pm 2(8) = 120 \pm 16$ ، أي أن الفترة من 104 إلى 136 درجة تضم على الأقل 75% (810 يوما) من المفردات.

9. الانحراف المعياري يعد أدق وأفضل مقياس ضمن مقاييس التشتت التي تضم المدى والمدى الربيعي والمئيني وحتى الانحراف المتوسط.

10. توجد علاقة استنباطية تقريبية بين بعض مقاييس التشتت مثل الانحراف المعياري والانحراف المتوسط ونصف المدى الربيعي كالتالي:

$$\mathrm{M.D}=\frac{4}{5}\,\mathrm{S.D}, \qquad \mathrm{SIQR}=\frac{2}{3}\,\mathrm{S.D}$$

[1] هذه النظرية لها استخدام آخر في تطبيق الاحتمالات كما سنرى لاحقا.

6.1.3 التباين (The Variance)

تعريف (7.3): **التباين:** التباين هو مربع الانحراف المعياري، ويرمز له بالرمز $Var(X)$ ويعطى بالصيغة:

$$Var(X) = (\mathrm{S.D})^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})^2}{N} = \frac{\sum_{i=1}^{N} x_i^2}{N} - \left(\frac{\sum_{i=1}^{N} x_i}{N}\right)^2$$

وهنا يجب الإشارة إلى أن التباين كمقياس للتشتت يستخدم أيضا لحساب درجة انتشار قيم البيانات حول وسطها الحسابي إلا أن الفرق بينه وبين الانحراف المعياري أن قيمة الأخير تمثل مقدار التشتت "بنفس" الوحدة التي تقاس بها مفردات البيانات (متر، درجة، يوم، ...) أما التباين فإن قيمته تمثل مقدار التشتت مربعا. لذلك فإن أكثر الباحثين ومستخدمي الأساليب الإحصائية يفضلون استخدام الانحراف المعياري كمقياس للتشتت لأن قيمته تعكس حقيقة الانتشار كما هي.

مثال (7.3): أوجد التباين مستخدما البيانات في المثال (4.3)، والتي تمثل أطوال 5 أنواع من الحشرات.

الحل:

$$Var(X) = (\mathrm{S.D})^2 = 2.96 \text{ سم}^2$$

وحيث أن طول الحشرات لا يقاس بالسنتيمتر المربع فيفضل التعليق على درجة تشتت أطوال الحشرات باستخدام الانحراف المعياري كما في المثال (4.3).

تعريف (8.3): **التباين المشترك** (Pooled Variance): يمكن حساب التباين المشترك لمتغيرين أو مجتمعين من البيانات بالصيغة التالية:

$$Var(X)_{\mathrm{Pooled}} = \frac{N_1 Var(X_1) + N_2 Var(X_2)}{N_1 + N_2}$$

حيث N_1 و N_2 هي عدد المفردات في المجموعة الأولى والثانية على الترتيب، و $Var(X_1)$ و $Var(X_2)$ هما تبايني المجموعة الأولى والثانية على الترتيب. ولاحظ أن الصيغة السابقة هي بمثابة متوسط حسابي موزون للتباينات. وعندما يكون $N_1 = N_2$ فإن:

$$Var(X)_{\mathrm{Pooled}} = \frac{Var(X_1) + Var(X_2)}{2}$$

مثال (8.3): أوجد التباين المشترك للبيانات في المثال (5.3).

الحل:

لدينا

$$Var(X_{\mathrm{A}}) = (\mathrm{S.D_A})^2 = (3.33)^2 = 11.09$$

وبالمثل

$$Var(X_{\mathrm{B}}) = (\mathrm{S.D_B})^2 = (6.43)^2 = 41.34$$

وحيث أن $N_1 = N_2 = 9$ فإن التباين المشترك لإنتاج المخبزين A و B هو

$$Var(X)_{\text{Pooled}} = \frac{11.09 + 41.34}{2} = 26.22$$

ويكون الانحراف المعياري المشترك لإنتاج المخبزين هو

$$\text{S.D}_{\text{Pooled}} = \sqrt{Var(X)_{\text{Pooled}}} = \sqrt{26.22} = 5.12 \text{ جرام}$$

ونلاحظ أن الزيادة في تشتت أوزان أرغفة الخبز في المخبز B قد "أثر" على نظيره في المخبز A فكان التشتت المشترك أقرب لتشتت أوزان الأرغفة في المخبز B.

7.1.3 معامل الاختلاف أو التشتت (Coefficient of Variation or Dispersion)

رأينا فيما سبق كيفية استخدام مقياس الانحراف المعياري للمقارنة بين درجات تشتت أو انتشار مفردات البيانات لمتغيرين (أو أكثر إذا أردنا) عندما يكون لهذه المتغيرات أوساط حسابية متقاربة، ولكن ماذا يحدث عندما تكون هذه المقارنة بين متغيرات لها أوساط مختلفة كثيرا في قيمتها و/أو لها وحدات قياس مختلفة، كأن نقارن بين درجات تشتت أسعار السلع الأساسية في عدة دول تستخدم عملات مختلفة، أو المقارنة بين قوة محركات مقاسة بقوة الحصان (HP) والكيلوات (kw). في هذه الحالة، نحن بحاجة لمقياس آخر يأخذ بعين الاعتبار هذا الاختلاف في وحدات القياس، وهو ما يدفعنا لاستخدام مقياس معامل الاختلاف أو التشتت والذي ينسب الانحراف المعياري لكل مجموعة من البيانات إلى وسطها الحسابي[1].

تعريف (9.3): **معامل الاختلاف:** معامل الاختلاف للمفردات $x_1, x_2, \dots, x_N$ هو النسبة بين الانحراف المعياري لهذه المفردات ووسطها الحسابي، ويرمز له بالرمز C.V :

$$\text{C.V} = \frac{\text{S.D}}{\bar{X}} \times 100 \quad , \bar{X} \neq 0$$

وكلما زادت قيمة معامل الاختلاف دل ذلك على زيادة درجة التشتت في البيانات.

مثال (9.3): مصنع لإنتاج الشاشات المسطحة للحواسيب يقوم بإنتاج نوعين من الشاشات LCD و LED. النوع الأول له معدل عمر استهلاكي يقدر بـ 1495 ساعة بانحراف معياري 280 ساعة، والنوع الثاني (LED) له معدل عمر استهلاكي يقدر بـ 1875 ساعة بانحراف معياري 310 ساعات. أي من نوعي الشاشات تعتقد أن له درجات تفاوت (اختلاف) أكبر في العمر الاستهلاكي؟.

الحل:

بما أن معدلي العمر الاستهلاكي لنوعي الشاشات مختلفان بدرجة كبيرة، نقوم باستخدام معامل الاختلاف للمقارنة:

---

[1] بعض المتخصصين قد يستخدم معامل الاختلاف لقياس درجة التشتت لمجموعة واحدة من البيانات أو لمتغير واحد بغرض مقارنته بمؤشر ما أو للمقارنة بين تشتت البيانات في فترات زمنية سابقة.

$$\mathrm{C.V}_1 = \frac{\mathrm{S.D}_1}{\bar{X}_1} \times 100 = \frac{280}{1495} \times 100 = 18.7\%$$

$$\mathrm{C.V}_2 = \frac{\mathrm{S.D}_2}{\bar{X}_2} \times 100 = \frac{310}{1875} \times 100 = 16.5\%$$

وهذا معناه أن نوع الشاشات الأول (LCD) له درجة تفاوت أو تشتت أكبر في العمر الاستهلاكي.

8.1.3 مقاييس التشتت لبيانات الجداول التكرارية
(Measurements of Dispersion for Tabulated Data)

والآن سنقوم بتعريف مقاييس التشتت التي تستخدم مع البيانات ذات المشاهدات المتكررة وبيانات جداول التوزيع التكراري.

تعريف (10.3): **مقاييس التشتت لبيانات الجداول التكرارية** (البيانات المبوبة): إذا كانت $x_1, x_2, \dots, x_k$ هي مفردات (أو قيم مراكز فترات لتوزيع تكراري) مناظرة للتكرارات $f_1, f_2, \dots, f_k$ على الترتيب، حيث $\sum_{i=1}^{k} f_i =$ N ، و k هي عدد الخلايا أو الفترات، فإن مقاييس التشتت يتم حسابها بالصيغ التالية:

1) المدى: يمكن حساب المدى بأكثر من طريقة عند التعامل مع البيانات في الجداول التكرارية، أشهرها:
   أ. المدى = الحد الأعلى الفعلي للفترة الأخيرة – الحد الأدنى الفعلي للفترة الأولى
   ب. المدى = مركز الفترة الأخيرة – مركز الفترة الأولى

ملاحظة: فيما يخص المدى الربيعي، نصف المدى الربيعي، والمدى المئيني فإن صيغ القوانين الخاصة بها لا تتغير لأن تطبيقها يعتمد فقط على حساب الربيعات والمئينات من الجدول التكراري.

2) الانحراف المتوسط:

$$\mathrm{M.D} = \frac{\sum_{i=1}^{k} f_i |x_i - \bar{X}|}{\sum_{i=1}^{k} f_i}$$

3) الانحراف المعياري:

$$\mathrm{S.D} = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \bar{X})^2}{\sum_{i=1}^{k} f_i}}$$

أو

$$= \sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i} - \left(\frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}\right)^2}$$

وعند التعامل مع العينات يكون

$$\mathrm{S.D} = \sqrt{\frac{n\sum_{i=1}^{k} f_i x_i^2 - \left(\sum_{i=1}^{k} f_i x_i\right)^2}{n(n-1)}}$$

حيث $n = \sum_{i=1}^{k} f_i$

ويكون التباين دائما هو مربع الانحراف المعياري بغض النظر عن كيفية حساب الأخير، وبالمثل يكون حساب معامل الاختلاف من الصيغة الأصلية.

4) **معامل الانحراف الربيعي** (Interquartile Deviation Coefficient):

عند التعامل مع الجداول التكرارية التي تضم فترات مفتوحة لا يمكن حساب الوسط الحسابي مباشرة، وبالتالي لا يمكن حساب معامل الاختلاف أيضا، لذلك يتم استخدام النسبة بين نصف المدى الربيعي[1] والوسيط بدلا من النسبة بين الانحراف المعياري والوسط الحسابي.

تعريف (11.2): **معامل الانحراف الربيعي:** يعرف معامل الانحراف الربيعي بالصورة:

$$C.V_Q = \frac{SIQR}{\tilde{X}} = \frac{Q_3 - Q_1}{Q_1 + Q_3}$$

حيث[2] SIQR هو نصف المدى الربيعي و $\tilde{X}$ هو الوسيط.

مثال (10.3): الجدول التالي (جدول (4.3)) يوضح عدد الساعات الأسبوعية التي يعتقد 70 طالبا من طلبة قسم الإحصاء أنها تكفيهم للمذاكرة في المنزل خلال العام الدراسي.

جدول (4.3): عدد ساعات الدراسة الأسبوعية لـ70 طالبا في قسم الإحصاء، (مثال (10.3)).

التكرار f (عدد الطلبة)	الفترة المستمرة (عدد الساعات)
20	2 – 4
25	4 – 6
15	6 – 8
5	8 – 10
5	10 – 12

والمطلوب حساب مقاييس التشتت التالية:

1. المدى
2. نصف المدى الربيعي
3. الانحراف المتوسط
4. الانحراف المعياري والتباين
5. على افتراض أن هؤلاء الطلبة هم نفسهم الذين تحصلوا على التقديرات الدراسية في المثال (25.2) في الفصل الثاني، فقارن بين درجة تشتت عدد ساعات دراستهم الأسبوعية وتقديراتهم الدراسية باستخدام معامل الاختلاف.

الحل:

1. المدى = الحد الأعلى الفعلي للفترة الأخيرة – الحد الأدنى الفعلي للفترة الأولى

المدى = 12 – 2 = 10 ساعات، أي أنه توجد تقريبا عشر ساعات اختلاف في الوقت المستغرق للدراسة بين الطلبة.

2. لحساب نصف المدى الربيعي نقوم أولا بإيجاد التكرار المتجمع الصاعد، (كما هو موضح في جدول (5.3))، وذلك لحساب الربيعان Q_1 و Q_3.

[1] لأن المدى أيضا لا يمكن حسابه في التوزيعات المفتوحة (الفترات المفتوحة).

[2] لاحظ أن $\frac{Q_3-Q_1}{Q_1+Q_3} = \frac{(Q_3-Q_1)/2}{(Q_1+Q_3)/2} = \frac{SIQR}{\tilde{X}}$.

جدول (5.3): التكرار المتجمع الصاعد للتوزيع التكراري في المثال (10.3).

الفترة المستمرة (عدد الساعات)	التكرار f (عدد الطلبة)	التكرار المتجمع الصاعد
2 – 4	20	20
4 – 6	25	45
6 – 8	15	60
8 – 10	5	65
10 – 12	5	70

رتبة Q_1 هي $\frac{N}{4}=\frac{70}{4}=17.5$ ، وبالتالي تكون فترة الربيع الأول هي $(4-2)$ وبالتالي

$$Q_1 = L_1 + \left[\frac{\left(\frac{N}{4}\right) - F_1}{f_Q}\right] \times L_Q = 2 + \left[\frac{17.5-0}{20}\right] \times 2 = 3.75 \text{ ساعة}$$

أي أن 25% من هؤلاء الطلبة يذاكرون لمدة تقل عن 3.75 ساعة في الأسبوع[1].

من جديد تكون رتبة Q_3 هي $\frac{3N}{4}=\frac{3\times 70}{4}=52.5$ ، فتكون فترة الربيع الثالث هي $(8-6)$ وبالتالي

$$Q_3 = L_1 + \left[\frac{\left(\frac{3N}{4}\right) - F_1}{f_Q}\right] \times L_Q = 6 + \left[\frac{52.5-45}{15}\right] \times 2 = 7 \text{ ساعة}$$

وهذا يعني أن 25% فقط من هؤلاء الطلبة يذاكرون لمدة تزيد عن 7 ساعات أسبوعيا[2].

وهكذا فإن

$$\text{SIQR} = \frac{Q_3 - Q_1}{2} = \frac{7-3.75}{2} = 1.63 \text{ ساعة}$$

وتدل هذه النتيجة على أن الطلبة يختلفون عن بعضهم البعض في عدد ساعات الدراسة بمدة تقارب الساعة والنصف تقريبا.

3. لحساب الانحراف المتوسط نستخدم الأعمدة في جدول (6.3)، حيث نقوم أولا بحساب الوسط الحسابي لعدد ساعات المذاكرة من الجدول التكراري.

جدول (6.3): الحسابات الخاصة بإيجاد الانحراف المطلق والمعياري في المثال (10.3).

الفترة المستمرة	التكرار f	مركز الفترة x	$f_i x_i$	$\lvert x_i - \bar{X}\rvert$	$f_i\lvert x_i - \bar{X}\rvert$	$f_i x_i^2$
2 – 4	20	3	60	2.57	51.43	180
4 – 6	25	5	125	0.57	14.29	625
6 – 8	15	7	105	1.43	21.43	735
8 – 10	5	9	45	3.43	17.14	405
10 – 12	5	11	55	5.43	27.14	605
المجموع	70		390		131.43	2550

$$\bar{X} = \frac{\sum_{i=1}^{5} f_i x_i}{\sum_{i=1}^{5} f_i} = \frac{390}{70} = 5.57 \text{ ساعة}$$

[1] لاحظ أن $F_1 = 0$ لأن فترة Q_1 هي أول فترة في الجدول التكراري.

[2] حيث أن 75% من الطلبة يذاكرون لمدة تقل عن 7 ساعات أسبوعيا.

أي أن الطلبة يدرسون بمعدل 5 ساعات ونصف تقريبا في الأسبوع، وهذه القيمة تقع فعليا ضمن الفترة الثانية (فترة المنوال) وهذا منطقي لأن الفترتين الأولى والثانية تضم أكبر عدد من الطلبة.
ويتم حساب الانحراف المتوسط بالصورة

$$M.D = \frac{\sum_{i=1}^{5} f_i |x_i - \bar{X}|}{\sum_{i=1}^{5} f_i} = \frac{131.43}{70} = 1.88 \text{ ساعة}$$

أي أنه يوجد اختلاف مقداره 1.88 ساعة ضمن عدد ساعات الدراسة للطلبة.

4. نستخدم الجدول (6.3) نفسه لحساب الانحراف المعياري

$$S.D = \sqrt{\frac{\sum_{i=1}^{5} f_i x_i^2}{\sum_{i=1}^{5} f_i} - \left(\frac{\sum_{i=1}^{5} f_i x_i}{\sum_{i=1}^{5} f_i}\right)^2} = \sqrt{\frac{2550}{70} - \left(\frac{390}{70}\right)^2}$$
$$= \sqrt{5.39} = 2.32 \text{ ساعة}$$

وهذا يعني أن الطلبة في هذا المثال تختلف عدد ساعات دراستهم الأسبوعية عن المعدل (5.57 ساعة) "بمتوسط" انتشار يساوي 2.32 ساعة، وتكون هذه القيمة هي الأكثر مصداقية بالنسبة لنا من مقاييس التشتت الأخرى.
ويكون التباين

$$Var(X) = (S.D)^2 = 5.39$$

وهذه القيمة هي، كما وضحنا سابقا، تمثل مربع اختلاف عدد ساعات دراسة الطلبة عن المعدل 5.57 ساعة دراسة أسبوعية.

5. معامل الاختلاف لعدد ساعات الدراسة هو

$$C.V_1 = \frac{S.D_1}{\bar{X}_1} \times 100 = \frac{2.32}{5.57} \times 100 = 41.66\%$$

نقوم بحساب معامل الاختلاف لتقديرات الطلبة في المثال (25.2) فنحصل على

$$C.V_2 = \frac{S.D_2}{\bar{X}_2} \times 100 = \frac{20.65}{51.43} \times 100 = 40.16\%$$

من قيم كلا من المعاملين نرى أن درجة تشتت تقديرات الطلبة هي مقاربة جدا لدرجة تشتت عدد ساعات دراستهم الأسبوعية، وهذا قد يدفعنا للقول بأنه إذا ما حاول الطلبة زيادة عدد ساعات دراستهم الأسبوعية فإن هذا قد يؤدي بدوره لزيادة تقديراتهم الدراسية[1].
ونلاحظ أن استخدام مقياس معامل الاختلاف كان ضروريا للمقارنة بيم المتغيرين (عدد ساعات الدراسة والتقديرات الدراسية) لأنه ليس من المنطق مقارنة 2.32 ساعة (كانحراف معياري) بـ20.65 درجة دراسية.

---

[1] هذه التفسير يمكن تناوله من وجهة نظر أخرى تندرج تحت مفهوم **الارتباط** (Correlation) في الإحصاء.

ملاحظة: في بعض جداول التوزيع التكراري، قد يتم تكوين الفترات بطريقة تفتقد للخبرة والدقة، وهذا قد يؤدي بدوره لظهور أخطاء (رقمية) في حساب تباين مفردات البيانات، في هذه الحالة ينصح باستخدام صيغة تصحيح تعرف **بتصحيح شبرد للتباين**.

تعريف (12.3): **تصحيح شبرد للتباين** (Sheppard's Correction for Variance): تعرف صيغة شبرد لتصحيح التباين بالصورة

$$Var_{sh}(X) = Var(X) - \frac{c^2}{12}$$

حيث c هي طول الفترة في جدول التوزيع التكراري و $Var(X)$ هو التباين الأصلي.

مثال (11.3): استخدم تصحيح شبرد لتصحيح التباين في المثال (10.3).

الحل:

$$Var_{sh}(X) = Var(X) - \frac{c^2}{12} = 5.39 - \frac{2^2}{12} = 5.06$$

وبالتالي فإن الانحراف المعياري المصحح يكون

$$\mathrm{S.D}_{sh} = \sqrt{Var_{sh}(X)} = \sqrt{5.06} = 2.25 \text{ ساعة}$$

ولاحظ أنه في مثالنا هذا لا يوجد فرق كبير بين قيمة الانحراف المعياري الأصلي (2.32 ساعة) والانحراف المعياري المصحح.

9.1.3 الدرجات المعيارية (Standard Units or Z-scores)

لنفرض أن مهندسين صناعيين، هما أحمد وعمر مثلا، تقدموا للعمل بأحد مصانع الحديد والصلب ضمن مجموعة من المتقدمين. وقام المصنع بإجراء اختبار عملي وآخر كتابي لجميع المتقدمين. فتحصل أحمد على 82 درجة في الاختبار العملي و89 درجة في الاختبار الكتابي، وتحصل عمر على 85 درجة في الاختبار العملي و87 درجة في الاختبار الكتابي، وهذه النتائج مُدرجة في جدول (7.3). والآن إذا ما أردنا المقارنة بين المهندسين أحمد وعمر في الاختبار العملي نجد أن أداء المهندس عمر كان أفضل بمجرد المقارنة بين الدرجتين، ونجد من ناحية أخرى أن أداء المهندس أحمد كان أفضل في الاختبار الكتابي.

جدول (7.3): نتائج المهندسين في للاختبار العملي والكتابي.

		الاختبار	
		عملي	كتابي
المهندس	أحمد	82	89
	عمر	85	87

ولكن ماذا لو أردنا المقارنة بين درجتي أحمد في كلا من الاختبار العملي والكتابي، هل نستطيع القول أن درجته في الاختبار الكتابي (89) كانت أفضل من درجته في الاختبار العملي (82)؟، وإذا كان المطلوب هو المقارنة بين أداء أحمد في الاختبار العملي وأداء عمر في الاختبار الكتابي، فهل نقول أن أداء عمر (87) كان أفضل من أداء أحمد (82)؟.

في الواقع، لا يمكن أن تتم كل هذه المقارنات بصورة مباشرة لأن طبيعة كل اختبار (العملي والكتابي) تختلف عن الأخرى، لهذا يجب أن نقارن درجات كل مهندس باعتبار (نسبة إلى) ما تحصل عليه باقي المتقدمين، فهذه الدرجات تحدد ترتيب كل متقدم، والمهندس الذي يتحصل على أفضل ترتيب (أعلى درجة) يكون هو الأفضل للوظيفة.

ومن أجل الحصول على ترتيب المتقدمين، يجب أن نأخذ بالاعتبار توزيع مفردات بياناتهم (درجاتهم) ويتم هذا بتحديد الوسط الحسابي والانحراف المعياري كما سنرى الآن.

تعريف (13.3): **الدرجات المعيارية:** إذا كانت x هي أي مفردة ضمن مجموعة مفردات وسطها الحسابي $\bar{X}$ وانحرافها المعياري S.D فإن الدرجة المعيارية لـ x يتم حسابها بالصيغة:

$$z = \frac{x - \bar{X}}{\mathrm{S.D}}$$

مثال (12.3): قارن بين أداء المهندس أحمد في الاختبار العملي والكتابي باستخدام البيانات في الجدول السابق (7.3)، علما بأن متوسط درجات المتقدمين في الاختبار العملي هو 68 درجة بانحراف معياري 8 درجات، و متوسط الدرجات في الاختبار الكتابي هو 80 درجة بانحراف معياري 6 درجات.

الحل:

درجة أحمد في الاختبار العملي هي $x_1 = 82$ ودرجته في الاختبار الكتابي هي $x_2 = 89$، فتكون الدرجة المعيارية للاختبار العملي هي

$$z_1 = \frac{x_1 - \bar{X}_1}{\mathrm{S.D}_1} = \frac{82 - 68}{8} = 1.75$$

والدرجة المعيارية للاختبار الكتابي هي

$$z_2 = \frac{x_2 - \bar{X}_2}{\mathrm{S.D}_2} = \frac{89 - 80}{6} = 1.50$$

وهذا يعني أنه بالرغم من أن درجة أحمد في الاختبار الكتابي (89) كانت أكبر من درجته في الاختبار العملي (82)، إلا أن الدرجة المعيارية للاختبار العملي كانت أفضل. وقد يكون السبب أن مستوى الاختبار العملي كان أصعب من الكتابي بدليل انخفاض المستوى العام (معدل الدرجات) للمتقدمين ($x_1 = 82$).

ويمكن أيضا استخدام الدرجات المعيارية للمقارنة بين أكثر من مفردتين من البيانات، وأيضا للمقارنة بين مفردات لمتغيرات مختلفة في وحدة القياس، كأن تتم المقارنة بين درجات حرارة مُقاسة بالدرجة المئوية وبالفهرنهايت، أو بين سرعات مقاسة بالكيلومتر/ساعة و متر/ساعة، وغيرها.

2.3 العزوم، الالتواء، والتفرطح (Moments, Skewness, and Kurtosis)

تناولنا في الفصل السابق والأجزاء السابقة في الفصل الحالي أهم المقاييس الإحصائية التي تستخدم في وصف نزعة البيانات إلى وسطها الحسابي وحساب مدى انتشارها حول هذا الوسط. وضمن هذا السياق، فإن

الجزء الحالي سيوضح المزيد من المقاييس التي تستخدم لتحديد اتجاه تركز توزيع البيانات، وكيف يمكن للقيم أن تقترب أو تبتعد عن الوسط، سواء كان هذا الابتعاد بالزيادة (إلى اليمين)، أو بالنقصان (إلى اليسار).

1.2.3 العزوم (Moments)

العزوم هو مقياس كمي يستخدم لحساب تباين أو "دوران" المفردات حول قيمة معينة، ويقصد بالدوران، (وهو مصطلح فيزيائي – رياضي)، إحصائيا ابتعاد قيم البيانات عن قيمة معينة قد تكون الصفر أو الوسط أو أي ثابت آخر. ويمكن التفكير في العزوم بأنها تمثل الحالة العامة لأهم مقياسين إحصائيين من مقاييس النزعة المركزية والتشتت وهما الوسط الحسابي والانحراف المعياري كما سنرى.

تعريف (14.3): **العزوم للبيانات المفردة:** إذا كانت $x_1, x_2, \dots, x_N$ هي مفردات مجموعة من البيانات التي لها الوسط الحسابي $\bar{X}$ ، فإن العزوم يمكن أن تقسم (حسب النقطة التي يحسب مقدار الابتعاد عنها) إلى القسمين التاليين:

1) **العزوم حول الصفر، (أو حول نقطة الأصل)** (Moments about zero):

يعرف العزم الرائي (r^{th} moment) حول الصفر بالصيغة

$$\acute{m}_r = \frac{\sum_{i=1}^{N} x_i^r}{N} \quad , r = 1,2,3, \dots$$

ويتم التعويض في قانون العزوم بقيم r حيث ($r = 1,2,3, \dots$) فنحصل على العزم الأول، الثاني، الثالث، ... وهكذا.

2) **العزوم حول الوسط الحسابي** (Moments about mean):

يعرف العزم الرائي حول الوسط الحسابي بالصيغة

$$m_r = \frac{\sum_{i=1}^{N} (x_i - \bar{X})^r}{N} \quad , r = 1,2,3, \dots$$

ملاحظات (حول العزوم):

1. عند استخدام أي قيمة ثابتة أخرى (A) مكان الوسط يكون العزم الرائي حول هذه القيمة هو

$$m_r = \frac{\sum_{i=1}^{N} (x_i - A)^r}{N} \quad , r = 1,2,3, \dots$$

2. إذا تم التعويض عن قيم r الأولى والثانية في قانون العزوم حول الصفر فإننا نحصل على

$$\acute{m}_1 = \frac{\sum_{i=1}^{N} x_i}{N} = \bar{X}$$

أي أن العزم الأول حول الصفر يساوي الوسط الحسابي للمفردات. وكذلك يكون

$$\acute{m}_2 = \frac{\sum_{i=1}^{N} x_i^2}{N}$$

وهو أحد المقدارين اللذين يستخدمان في حساب التباين، ومنها نستطيع كتابة

$$Var(X) = \frac{\sum_{i=1}^{N} x_i^2}{N} - \left(\frac{\sum_{i=1}^{N} x_i}{N}\right)^2 = \acute{m}_2 - \acute{m}_1^2$$

3. وكذلك إذا ما تم التعويض عن القيم $r = 1,2$ في قانون العزوم حول الوسط الحسابي فإننا نحصل على

$$m_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})}{N} = 0$$

حيث أن مجموع انحرافات قيم المفردات عن وسطها الحسابي يساوي الصفر. وأيضا

$$m_2 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})^2}{N} = Var(X)$$

أي أن العزم الثاني حول الوسط الحسابي هو تباين المفردات.

مثال (13.3): أوجد العزوم الثلاثة الأولى (أ) حول الصفر و (ب) حول الوسط الحسابي، وذلك للبيانات التالية: 10، 8، 7، 3، 2 . ثم أوجد الوسط الحسابي والانحراف المعياري للبيانات باستخدام العزوم.

الحل:

(أ)

$$\acute{m}_1 = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{30}{5} = 6$$

$$\acute{m}_2 = \frac{\sum_{i=1}^{5} x_i^2}{5} = \frac{2^2 + 3^2 + 7^2 + 8^2 + 10^2}{5} = \frac{226}{5} = 45.2$$

$$\acute{m}_3 = \frac{\sum_{i=1}^{5} x_i^3}{5} = \frac{2^3 + 3^3 + 7^3 + 8^3 + 10^3}{5} = \frac{1890}{5} = 378$$

ويكون

$$\bar{X} = \acute{m}_1 = 6$$

و

$$\text{S.D} = \sqrt{\acute{m}_2 - \acute{m}_1^2} = \sqrt{45.2 - 6^2} = \sqrt{9.2} = 3.03$$

(ب)

$$m_1 = \frac{\sum_{i=1}^{5}(x_i - 6)^r}{5} = \frac{0}{5} = 0$$

$$m_2 = \frac{\sum_{i=1}^{5}(x_i - 6)^2}{5} = \frac{(2-6)^2 + (3-6)^2 + (7-6)^2 + (8-6)^2 + (10-6)^2}{5}$$
$$= \frac{46}{5} = 9.2$$

$$m_3 = \frac{\sum_{i=1}^{5}(x_i - 6)^3}{5} = \frac{(2-6)^3 + (3-6)^3 + (7-6)^3 + (8-6)^3 + (10-6)^3}{5}$$
$$= \frac{-18}{5} = -3.6$$

و $\text{S.D} = \sqrt{m_2} = \sqrt{9.2} = 3.03$.

أما عند التعامل مع البيانات المبوبة (ذات المشاهدات المتكررة أو بيانات جداول التوزيع التكراري) فإن صيغ العزوم تكون كما هو موضح في التعريف التالي:

تعريف (15.3): **العزوم للبيانات المبوبة:** إذا كانت $x_1, x_2, \dots , x_k$ هي مشاهدات (أو مراكز فترات) لها التكرارات المناظرة $f_1, f_2, \dots, f_k$، ولها الوسط الحسابي $\bar{X}$ ، فإنه يمكن حساب العزوم حول الصفر وحول الوسط الحسابي بالصورة:

1) العزوم حول الصفر:

$$\acute{m}_r = \frac{\sum_{i=1}^{k} f_i x_i^r}{\sum_{i=1}^{k} f_i} \quad , r = 1,2,3, \dots$$

2) العزوم حول الوسط الحسابي:

$$m_r = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{X})^r}{\sum_{i=1}^{k} f_i} \quad , r = 1,2,3, \dots$$

ملاحظة: العلاقة بين العزوم حول الصفر والعزوم حول الوسط الحسابي، (عندما $r = 2,3,4$)، تكون بالصورة:

$$m_2 = \acute{m}_2 - \acute{m}_1^2$$

$$m_3 = \acute{m}_3 - 3\acute{m}_1\acute{m}_2 + 2\acute{m}_1^3$$

$$m_4 = \acute{m}_4 - 4\acute{m}_1\acute{m}_3 + 6\acute{m}_1^2\acute{m}_2 - 3\acute{m}_1^4$$

وهذه العلاقات يمكن اشتقاقها باستخدام مفكوك ذو الحدين[1].

مثال (14.3): أوجد العزوم الأربعة الأولى (أ) حول الصفر ، و (ب) حول الوسط الحسابي ، وذلك للمشاهدات في جدول التوزيع التكراري التالي:

الحل:

(أ)

لتبسيط الحسابات نقوم بتكوين أعمدة جديدة مستخدمين جدول (8.3) تتضمن المجاميع المطلوبة لحساب العزوم حول الصفر فيتكون لدينا الجدول (9.3).

جدول (8.3): التوزيع التكراري للمشاهدات في المثال (14.3).

التكرار f	الفترة المستمرة
3	1 − 3
5	3 − 5
4	5 − 7
2	7 − 9

[1] سيتم التطرق لمفكوك ذي الحدين (Binomial Expansion) في الفصل الرابع (الجزء (1.3.4))، إلا أنه يمكن حاليا استخدام الصيغة التالية: إذا كان X و Y متغيران يأخذان قيم حقيقية و كان $n > 0$ ، فإن:

$$(X+Y)^n = \frac{n!}{(n-0)!0!} X^n Y^0 + \frac{n!}{(n-1)!1!} X^{n-1} Y^1 + \dots + \frac{n!}{(n-n)!n!} X^0 Y^n$$

جدول (9.3): الحسابات الخاصة بإيجاد العزوم حول الصفر في المثال (14.3).

الفترة المستمرة	التكرار f	مركز الفترة x	f_ix_i	$f_ix_i^2$	$f_ix_i^3$	$f_ix_i^4$
1 − 3	3	2	6	12	24	48
3 − 5	5	4	20	80	320	1280
5 − 7	4	6	24	144	864	5184
7 − 9	2	8	16	128	1024	8192
المجموع	14		66	364	2232	14704

وهكذا فإن

$$\acute{m}_1 = \frac{\sum_{i=1}^{4} f_i x_i}{14} = \frac{66}{14} = 4.71 = \bar{X}$$

$$\acute{m}_2 = \frac{\sum_{i=1}^{4} f_i x_i^2}{14} = \frac{364}{14} = 26$$

$$\acute{m}_3 = \frac{\sum_{i=1}^{4} f_i x_i^3}{14} = \frac{2232}{14} = 159.43$$

$$\acute{m}_4 = \frac{\sum_{i=1}^{4} f_i x_i^4}{14} = \frac{14704}{14} = 1050.29$$

(ب)

بالمثل نقوم بتكوين أعمدة جديدة في جدول (8.3) لحساب المجاميع الخاصة بحساب العزوم حول الوسط الحسابي فنحصل على جدول (10.3)، علما بأن الوسط الحسابي قد تم حسابه في المطلوب (أ).

جدول (10.3): الحسابات الخاصة بإيجاد العزوم حول الوسط الحسابي في المثال (14.3).

الفترة المستمرة	التكرار f	مركز الفترة x	$f_i(x_i-\bar{X})$	$f_i(x_i-\bar{X})^2$	$f_i(x_i-\bar{X})^3$	$f_i(x_i-\bar{X})^4$
1 − 3	3	2	−8.14	22.10	−59.99	162.83
3 − 5	5	4	−3.57	2.55	−1.82	1.30
5 − 7	4	6	5.14	6.61	8.50	10.93
7 − 9	2	8	6.57	21.59	70.94	233.10
المجموع	14		0	52.86	17.63	408.17

وبالتالي فإن

$$m_1 = 0$$

$$m_2 = \frac{\sum_{i=1}^{4} f_i (x_i - \bar{X})^2}{14} = \frac{52.86}{14} = 3.78$$

$$m_3 = \frac{\sum_{i=1}^{4} f_i (x_i - \bar{X})^3}{14} = \frac{17.63}{14} = 1.26$$

$$m_4 = \frac{\sum_{i=1}^{4} f_i (x_i - \bar{X})^4}{14} = \frac{408.17}{14} = 29.15$$

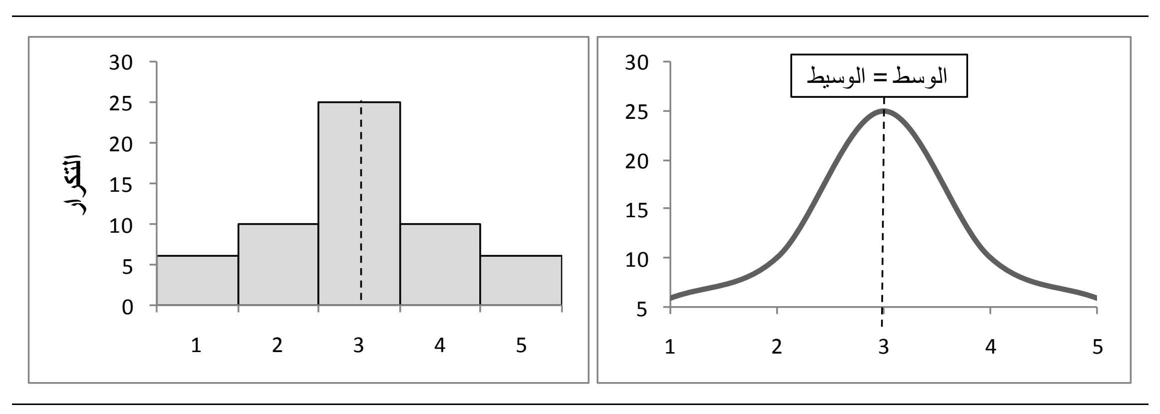
وبالطبع يكون من الأسهل عمليا أن يتم إيجاد العزوم الأربعة الأولى حول الوسط الحسابي من خلال استخدام علاقاتها بالعزوم حول الصفر.

2.2.3 الالتواء والتفرطح (Skewness and Kurtosis)

ناقشنا في هذا الفصل كيفية استخدام مقاييس النزعة المركزية والتعبير عنها أو تمثيلها بيانيا، ووضحنا أيضا أهمية التمثيل البياني في فهم توزيع البيانات. وفي الواقع، كثير من الأساليب الإحصائية الرئيسية في مجال الإحصاء الاستدلالي تعتمد على شكل توزيع البيانات اعتمادا كبيرا، ومن أمثلة ذلك اعتماد معظم الأساليب الإحصائية على افتراض أن توزيع البيانات هو توزيع طبيعي[1]. وتوزع البيانات طبيعيا هو أهم حالات ما يعرف بالتوزيعات المتماثلة، وسنتطرق فيما يلي لمفهوم توزع البيانات بشكل متماثل بهدف توضيح معنى الالتواء والتفرطح.

تعريف (16.3): **التماثل** (Symmetry): يقال عن توزيع البيانات أنه توزيع متماثل إذا ما انطبق طرفي (نصفي) التوزيع على بعضهما البعض على المحور الأفقي.

وشكل توزيع البيانات يمكن تمثيله بيانيا من خلال استخدام المدرج التكراري أو المنحنى التكراري أو كلاهما. ففي الشكل (3.3) يتضح لنا شكل التماثل حيث أننا إذا ما افترضنا أنه تم رسم خط مستقيم عمودي (الخط المنقط) في منتصف توزيع البيانات فإن طرفي التوزيع سينطبقان على بعضهما البعض تماما.



شكل (3.3): توزيع متماثل لبيانات افتراضية.

وفي التوزيع المتماثل يكون عدد مفردات البيانات التي قيمها أقل من الوسط الحسابي مساوية لعدد المفردات التي قيمها أكبر منه، وهذا يعني أنه في البيانات ذات التوزيع المتماثل تكون قيمة الوسط الحسابي مساوية لقيمة الوسيط كما يرى في شكل (3.3).

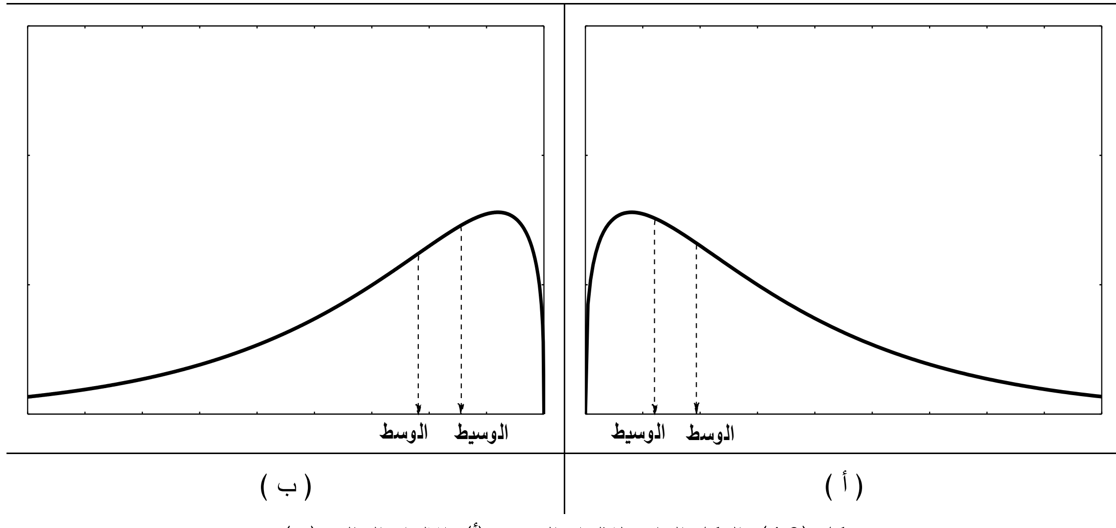
وبناءا على تعريف التماثل، فإن أي توزيع بيانات يأخذ شكلا آخر مختلف عن الشكل المتماثل يسمى توزيع غير متماثل أو توزيع ملتوي.

[1] سنتعرض لمفهوم التوزيع الطبيعي بشكل موسع عند دراسة التوزيعات الاحتمالية في الفصول القادمة.

تعريف (17.3): **الالتواء** (Skewness): يقال بأن توزيع البيانات هو توزيع ملتو إذا ما تجمعت مفردات البيانات في أحد طرفي التوزيع بصورة أكبر من الطرف الآخر.

وهنالك نوعان من الالتواء بحسب طبيعة تجمع مفردات البيانات. النوع الأول هو **الالتواء الموجب** (Positive Skewness) أو الالتواء نحو اليمين[1]، وهو الذي يظهر عند تجمع أو تكتل القيم (ممثلة بقمة المنحنى) نحو يسار الشكل البياني كما يوضح الشكل (4.3– أ)، وتكون "أكثر" المفردات في هذه الحالة قيمها أقل من قيمة الوسط الحسابي، والذي في نفس الوقت تكون قيمته أكبر من قيمة الوسيط.

أما النوع الثاني، والذي يظهر في الشكل (4.3– ب)، فتتجمع فيه معظم قيم المفردات ناحية اليمين ويسمى **بالالتواء السالب** (Negative Skewness) أو الالتواء نحو اليسار، وبعكس النوع الأول تكون قيمة الوسط الحسابي أقل من قيمة الوسيط. وفي كلا النوعين، تزيد حده الالتواء بزيادة تجمع مفردات البيانات في أحد الطرفين.



شكل (4.3): الشكل البياني للالتواء الموجب (أ) والالتواء السالب (ب).

ملاحظة: في التوزيعات التي يكون لها منوال وحيد يكون:

1) الوسط = الوسيط = المنوال ، عندما يكون توزيع البيانات متماثل.
2) الوسط > الوسيط > المنوال ، في التوزيع الملتوي الموجب.
3) الوسط < الوسيط < المنوال ، في التوزيع الملتوي السالب.

إضافة إلى مراقبة توزيع البيانات من خلال التمثيل البياني، فإنه يمكن حساب درجة ونوع الالتواء من خلال استخدام بعض المقاييس الرياضية التالية:

تعريف (18.3): **مقاييس الالتواء:** إذا كان $x_1, x_2, \dots, x_N$ هي مجموعة من المفردات، وكانت $\bar{X}$، $\tilde{X}$، $\hat{X}$، و $S.D(X)$ هي على الترتيب الوسط الحسابي، الوسيط، المنوال، والانحراف المعياري لهذه المفردات. فإنه يمكن قياس درجة ونوع الالتواء لتوزيع البيانات باستخدام الصيغ التالية:

[1] يقصد بذلك امتداد ذيل المنحنى الأطول ناحية يمين الشكل.

1. **معامل الالتواء المنوالي** أو **معامل بيرسون الأول للالتواء** (Pearson's First Coefficient of Skewness):

$$S.K_{\hat{X}} = \frac{\bar{X} - \hat{X}}{\mathrm{S.D}(X)}$$

2. **معامل الالتواء الوسيطي** أو **معامل بيرسون الثاني للالتواء** (Pearson's Second Coefficient of Skewness):

$$S.K_{\tilde{X}} = \frac{3(\bar{X} - \tilde{X})}{\mathrm{S.D}(X)}$$

3. **معامل الالتواء العزمي** (Moment Coefficient of Skewness):

$$S.K_m = \frac{m_3^2}{(\mathrm{S.D}(X))^3}$$

حيث m_3 هو العزم الثالث حول الوسط الحسابي.

4. **معامل الالتواء الربيعي** (Quartile Coefficient of Skewness):

$$S.K_{\mathrm{Q}} = \frac{\mathrm{Q}_3 - 2\mathrm{Q}_2 + \mathrm{Q}_1}{\mathrm{Q}_3 - \mathrm{Q}_1}$$

حيث Q_1، Q_2، و Q_3 هي الربيعات الثلاثة للبيانات.

5. **معامل الالتواء المئيني** (10-90 Percentile Coefficient of Skewness):

$$S.K_{\mathrm{P}} = \frac{\mathrm{P}_{90} - 2\mathrm{P}_{50} + \mathrm{P}_{10}}{\mathrm{P}_{90} - \mathrm{P}_{10}}$$

حيث P_{10}، P_{50}، و P_{90} هي المئين العاشر، الخمسون، و التسعون على الترتيب للبيانات.

ويكون نوع الالتواء موجب، (التواء ناحية اليمين)، إذا ما كانت الإشارة الجبرية لمعامل الالتواء موجبة والعكس بالعكس، وذلك لكل المعاملات السابقة[1]. أما عندما تكون قيمة معامل الالتواء مساوية للصفر فهذا يعني أن توزيع البيانات متماثل، وكلما زادت القيمة (المطلقة) للمعامل زادت درجة الالتواء، أي زاد عدد المفردات المتجمعة في أحد الأطراف.

مثال (15.3): البيانات التالية، (جدول (11.3))، تمثل درجات الحرارة المئوية المسجلة في إحدى المدن الأوروبية لمدة ثلاثين يوما خلال فصل الشتاء. والمطلوب وصف توزيع البيانات باستخدام مفهوم الالتواء كمقياس رياضي وبصورة بيانية أيضا.

جدول (11.3): توزيع درجات الحرارة في المثال (15.3).

الفترة	2.5−5.5	5.5−8.5	8.5−11.5	11.5−14.5	14.5−17.5	17.5−20.5
التكرار	2	3	6	6	8	5

[1] بالنسبة لمعامل بيرسون لالتواء فإن قيمته ستتراوح ما بين −3 و 3 .

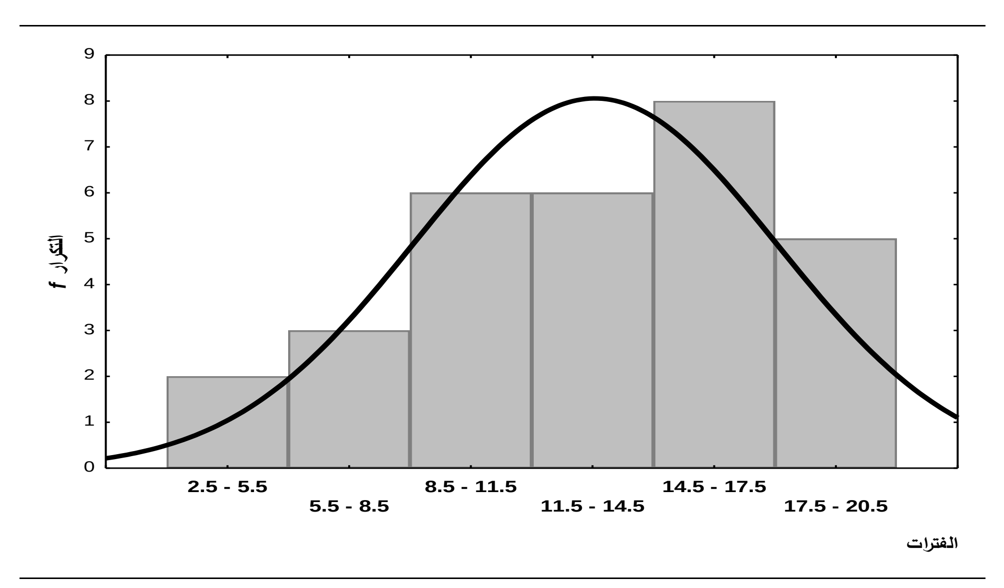
الحل:

نستطيع استخدام أي من المعاملات في تعريف (18.3)، ولنأخذ معامل الالتواء الوسيطي مثلا، فنكون بحاجة لحساب الوسط الحسابي، الوسيط، والانحراف المعياري للبيانات؛

لدينا $\bar{X} = 13$ ، $\tilde{X} = 13.5$ ، و $\mathrm{S.D}(X) = 4.38$ ، فيكون

$$S.K_{\tilde{X}} = \frac{3(\bar{X} - \tilde{X})}{\mathrm{S.D}(X)} = \frac{3(13 - 13.5)}{4.38} = -0.34$$

وهذا يشير إلى وجود التواء بسيط سالب. ويمكن ملاحظة ذلك أيضا في جدول (11.3) حيث نجد أن تكرار المفردات في الفترات الثلاثة الأخيرة هو أعلى منه في الفترات الثلاثة الأولى. وهذا يعني أن درجات الحرارة ما بين 11.5° و 20.5° كانت هي السائدة خلال ذلك الشهر، وهذه النتيجة تتضح أيضا من خلال النظر للمدرج التكراري الخاص بهذه البيانات في الشكل (5.3)، حيث يمكن مشاهدة تجمع المفردات (قمة المنحنى) إلى اليمين مما يؤدي لتكوين الالتواء ناحية اليسار، أي التواء سالب.



شكل (5.3): المدرج التكراري الخاص بدرجات الحرارة للمثال (15.3).

إضافة إلى ما سبق، فإن مراقبة درجة ونوع الالتواء على الرسم البياني، (للمدرج أو المنحنى التكراري) يمكن أن تكون مفيدة في التعرف على القيم المتطرفة ضمن البيانات، فكلما كان الطرف (الذيل) الأيمن للمنحنى أو المدرج أكثر امتدادا دل ذلك على وجود قيم متطرفة عليا ضمن المفردات، وبالعكس كلما كان الطرف الأيسر للمنحنى أكثر طولا كلما زاد ابتعاد أو تطرف بعض قيم البيانات الدنيا.

لقد رأينا أنه عند مراقبة درجة ونوع الالتواء في توزيع البيانات، فإن التركيز يكون على تجمع مفردات البيانات يمينا أو يسارا، ولكن هنالك نوع آخر من التغير في شكل التوزيع يمكن مراقبته أيضا وهو ارتفاع وانخفاض قمة المنحنى أو المدرج التكراري، وهو ما يعرف بمراقبة تفرطح توزيع البيانات.

تعريف (19.3): **التفرطح** (Kurtosis): يعرف التفرطح بأنه درجة "تدبب" أو تسطح توزيع البيانات. فإذا كانت مفردات البيانات تتركز قرب المركز وبعيدا عن الأطراف، (مما يؤدي لتدبب قمة منحنى التوزيع)، فعندها يسمى

توزيع البيانات **بالتوزيع المدبب** (Leptokurtic)، وإذا كانت المفردات مبتعدة عن القمة باتجاه الأطراف، (مما يؤدي لتسطح قمة المنحنى)، فإن التوزيع يكون **توزيعا مفرطحا** (Platykurtic). أما إذا كان توزيع البيانات مابين النوعين السابقين بحيث لا يظهر فيه "حدة" في انتشار البيانات فإن توزيع البيانات عندئذ يكون **توزيعا معتدلا** (Mesokurtic).

والشكل (6.3) يوضح لنا الصور الثلاث الأساسية لتفرطح توزيع البيانات بصورة عامة.



شكل (6.3): الأشكال الثلاثة الرئيسية لتفرطح البيانات.

وكما هو الحال مع الالتواء، فإنه يمكن حساب درجة ونوع التفرطح من خلال استخدام المقاييس الرياضية التالية:

تعريف (20.3): **مقاييس التفرطح:** إذا كان $x_1, x_2, \ldots, x_N$ هي مجموعة من المفردات، فإنه يمكن قياس درجة تفرطح توزيع هذه المفردات بالصيغ التالية:

1. **معامل التفرطح العزمي** (Moment Coefficient of Kurtosis):

$$Kur_m = \frac{m_4}{[\mathrm{S.D}(X)]^4}$$

حيث m_4 هو العزم الرابع للمفردات حول الوسط الحسابي و $\mathrm{S.D}(X)$ هو الانحراف المعياري للمفردات. ويكون التعليق على قيم المعامل بالصورة التالية:

إذا كان	$Kur_m > 3$	يكون توزيع البيانات	مدببا
	$Kur_m = 3$		معتدلا
	$Kur_m < 3$		مفرطحا

2. **معامل التفرطح المئيني** ((Percentile Coefficient of Kurtosis (10 – 90)):

$$Kur_{\mathrm{P}} = \frac{\frac{1}{2}(\mathrm{Q}_3 - \mathrm{Q}_1)}{\mathrm{P}_{90} - \mathrm{P}_{10}}$$

حيث Q_1 و Q_3 هما الربيعان الأول والثالث للمفردات و P_{10} و P_{90} هما المئينان العاشر والتسعون للمفردات.

ونستخدم المعيار التالي للتعليق على قيم المعامل:

إذا كان	$Kur_{\text{P}} < 0.263$	يكون توزيع البيانات	مدببا
	$Kur_{\text{P}} = 0.263$		معتدلا
	$Kur_{\text{P}} > 0.263$		مفرطحا

مثال (16.3): مستخدما البيانات في جدول (11.3)، (المثال (15.3))، أعط وصفا لتفرطح توزيع البيانات مستخدما معاملي الالتواء العزمي والمئيني.

الحل:

لدينا $m_4 = 934.9$ ، $\text{Q}_1 = 9.75$ ، $\text{Q}_3 = 16.56$ ، $\text{S.D}(X) = 4.38$ ، $\text{P}_{10} = 6.5$ ، $\text{P}_{90} = 18.7$ ، وبالتالي فإن

$$Kur_m = \frac{m_4}{[\text{S.D}(X)]^4} = \frac{934.9}{(4.38)^4} = 2.54 < 3$$

و

$$Kur_{\text{P}} = \frac{\frac{1}{2}(\text{Q}_3 - \text{Q}_1)}{\text{P}_{90} - \text{P}_{10}} = \frac{\frac{1}{2}(16.56 - 9.75)}{18.7 - 6.5} = 0.275 > 0.263$$

مما يدل على أن توزيع البيانات هو توزيع أقرب إلى الاعتدال في ارتفاع منحنى البيانات، وبالنظر لجدول (11.3) نجد أن سبب في هذا هو اقتراب القيم الكبيرة للتكرارات في الفترات التي في المنتصف (6، 6، 8) من بعضها البعض، وهذا يعني أن توزيع هذه البيانات هو ملتو وغير مفرطح كثيرا.

ملاحظات:

1. في التوزيعات المفرطحة، التي تمتد فيها أطراف المنحنى التكراري كثيرا، تكون قيمة الانحراف المعياري كبيرة لأن مفردات البيانات تنتشر (تتشتت) بصورة أكبر مما هو عليه في التوزيعات المدببة قصيرة الأطراف.
2. استخدام معاملات الالتواء والتفرطح يكون مفيدا في وصف شكل توزيع البيانات ودراسة سلوك التغير ضمن قيم المفردات، إلا أنه يستخدم أيضا للمقارنة بين توزيعين أو أكثر أو للمقارنة بين التغير في قيم البيانات بعد مرور فترة زمنية محددة، وذلك للمساعدة في دراسة الأسباب المؤدية للتغير في سلوك هذه المفردات.

3.3 بعض الرسومات البيانية الإضافية (Some Additional Graphical Displays)

في الأجزاء السابقة ضمن الفصل الحالي والسابق، تناولنا طرق استخدام بعض المقاييس الإحصائية الهامة في استكشاف ووصف مجموعة أو أكثر من البيانات، ووضحنا أهمية وسبب استخدام هذه المقاييس طبقا لطبيعة البيانات ونوع الدراسة المطلوب إجراؤها.

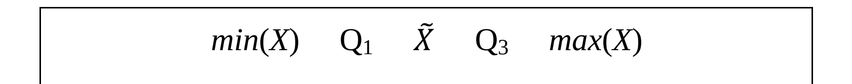
في هذا الجزء، سنقوم بعرض نوعين إضافيين من الرسومات البيانية التي كثيرا ما يتم إهمالها من قبل بعض الباحثين ومستخدمي الأساليب الإحصائية عند إجراء الدراسات الوصفية، رغم أهمية هذه الرسومات وما تقدمه من وصف واضح ومبسط لسلوك البيانات.

1.3.3 شكل الصندوق (The Boxplot)

وقد تم التنويه عليه في الجزء الخاص بالرسومات البيانية في الفصل الثاني، وهو ببساطة رسم بياني يتم من خلاله توضيح خمسة مقاييس هامة تمثل بحد ذاتها ملخصا، قد يكون بمفرده كافيا في بعض الأحيان، لتوضيح لسلوك مفردات البيانات.

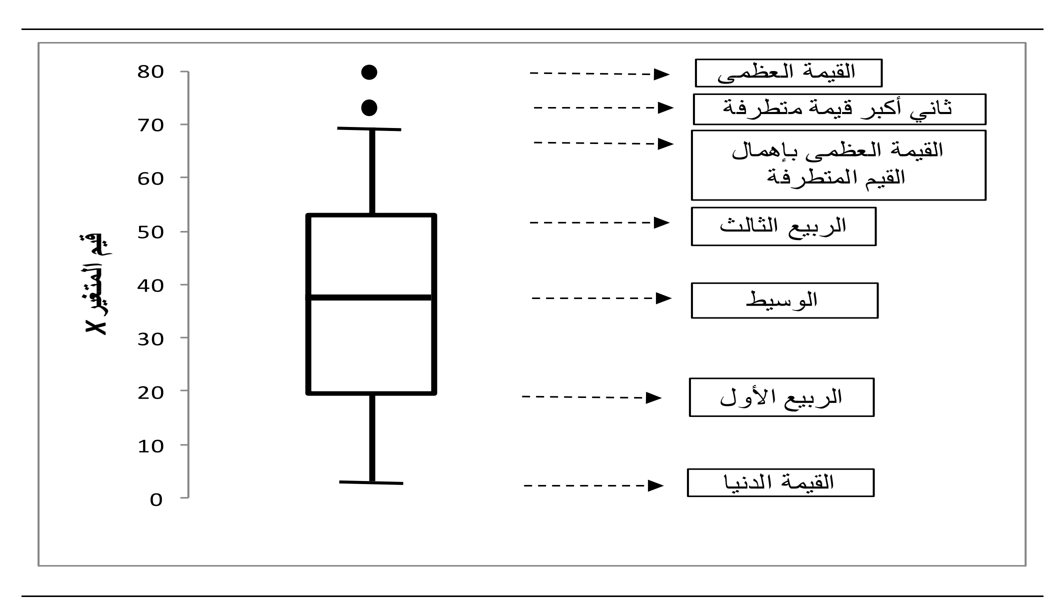
وهذا الرسم يمكن استخدامه لوصف مجموعة واحدة (متغير واحدة) من البيانات أو للمقارنة بين متغيرين أو أكثر، إذ يمكن من خلاله مراقبة تشتت توزيع البيانات عن مركزها وكذلك ملاحظة التوائها وتفرطحها، وأيضا يمكننا الصندوق من التعرف على القيم المتطرفة بصورة واضحة جدا كما سنرى.

إن المقاييس الخمسة التي يعرضها الصندوق، والتي تسمى أيضا **بملخص المقاييس أو الأرقام الخمسة**[1] (Five-Numbers Summary)، هي القيمة الدنيا للبيانات ((x_{min}) أو $(min(X))$) القيمة العظمى للبيانات ((x_{max}) أو $(max(X))$)، الربيع الأول (Q_1)، الربيع الثالث (Q_3)، والوسيط ($\tilde{X}$). وعادة ما يتم عرضها أفقيا بالصورة التالية:



$min(X) \quad Q_1 \quad \tilde{X} \quad Q_3 \quad max(X)$

ويكون الشكل التقليدي لرسم الصندوق على الصورة التالية، (شكل (7.3))؛



شكل (7.3): شكل بياني افتراضي لرسم الصندوق موضحا عليه المقاييس الإحصائية المرتبطة به.

[1] هذه المقاييس الخمسة عادة ما توفرها البرامج الإحصائية الشهيرة، (مثل S-plus ، R ، و Statistica)، بصورة تلقائية عند طلب ملخص إحصائي سريع للبيانات.

ويلاحظ من الشكل (7.3) أنه إذا لم توجد قيم متطرفة في توزيع البيانات، فإن القيمة العظمى ستكون هي القيمة عند الخط الذي يمثل الحد الأعلى في الصندوق، وكذلك الحال بالنسبة للقيمة الدنيا في البيانات. ولتوضيح كيفية استخدام رسم الصندوق في المقارنة، لنأخذ المثال التالي:

مثال (17.3): البيانات التالية (جدول (12.3)) تمثل مستوى الدخل الشهري (بالدولار الأمريكي) لمجموعتين من الأفراد[1] في دولتين مختلفتين، والمطلوب استخدام رسم الصندوق للمقارنة "الاستكشافية" بين مستوى الدخل في الدولتين.

جدول (12.3): مستوى الدخل الشهري لأفراد في الدولتين، (مثال (17.3)).

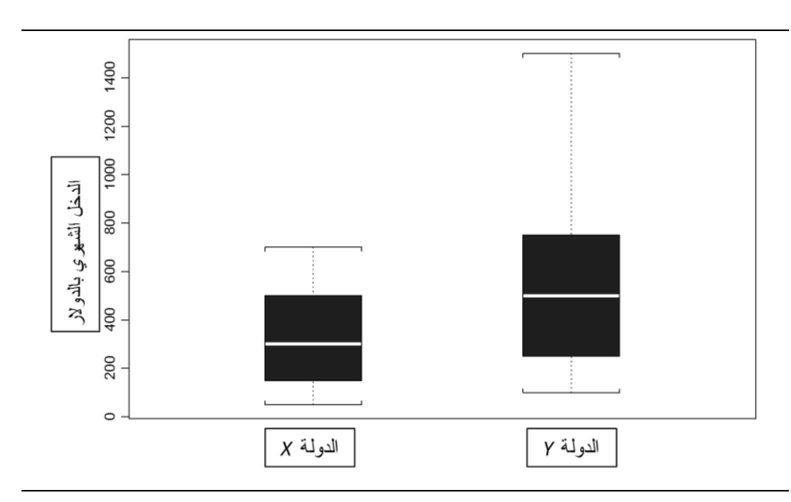
الدولة X	500	350	250	700	300	600	100	450	50	150	200
الدولة Y	1500	750	500	800	450	100	300	700	250	600	150

الحل:

نبدأ بإيجاد ملخص الأرقام الخمسة للمتغيرين X و Y بالصورة التالية:

المتغير (الدولة) Y					المتغير (الدولة) X				
$max(Y)$	Q_3	$\tilde{Y}$	Q_1	$min(Y)$	$max(X)$	Q_3	$\tilde{X}$	Q_1	$min(X)$
1500	725	500	275	100	700	475	300	175	50

ويكون شكل الصندوق بالصورة:



شكل (8.3): رسم الصندوق الخاص بمتغيرات المثال (17.3).

من الشكل (8.3) نلاحظ أولا أن مستوى الدخل في الدولة Y هو بصورة عامة أعلى منه في الدولة X وذلك بسبب ارتفاع حجم "الصندوق"[2] الثاني (إلى اليمين) أكثر من الأول (إلى اليسار)، ولأن $\tilde{X} = 300 < \tilde{Y} = 500$.

---

[1] على افتراض أن هؤلاء الأفراد يعملون في مهن متجانسة في كلا الدولتين.

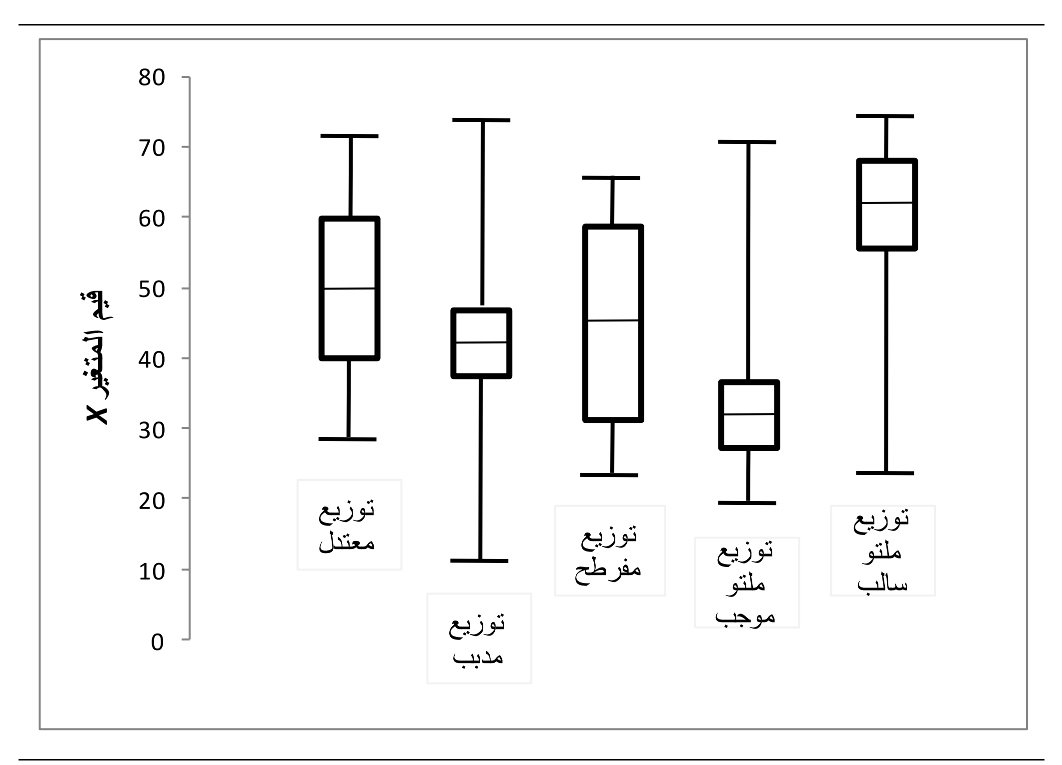
[2] حيث أن الصندوق في الشكل يمثل كتلة البيانات الأقرب للمركز (الوسيط).

ثانيا، نرى أن انتشار قيم الدخل الشهري في الدولة Y هو أكثر منه في الدولة X حيث أن الصندوق الثاني هو أكثر استطالة.

من جديد نلاحظ أن توزيع الدخل في الدولة Y هو أكثر التواء لأن تجمع قيم الدخل في شكل الصندوق الخاص بها هو أكبر، وكلا المتغيرين X و Y له التواء موجب (إلى اليمين)[1].

إضافة إلى ما سبق، فإن توزيع الدخل في الدولة Y يعتبر أكثر تفرطحا نظرا لامتداد حجم الصندوق الخاص به. ويلاحظ أيضا وجود قيمة متطرفة (نسبيا) ضمن قيم الدخل في الدولة Y وهي القيمة 1500 والتي تظهر بشكل واضح في أعلى الرسم الخاص بالدولة Y.

ملاحظة: يمكننا وضع نمط قياسي للمقارنة بين درجات التواء وتفرطح توزيع البيانات باستخدام رسم الصندوق كما هو موضح في شكل (9.3).



شكل (9.3): مراقبة درجات الالتواء والتفرطح في توزيع البيانات عند استخدام شكل الصندوق.

2.3.3 شكل الساق والورقة (The Stem-leaf plot)

وهو نوع من الرسومات البيانية التي تستخدم قيم البيانات نفسها في توضيح توزيع البيانات، وعادة ما يستخدم عندما يكون حجم البيانات ضئيل نسبيا. ويتم تطبيقه باستخدام الخطوات التالية:

- يتم رسم عمودين، الأول (إلى اليسار) ويمثل الساق، والثاني (إلى اليمين) ويمثل الورقة.

---

[1] لأن قيم الدخل تتجمع نحو القيم الأقل، وهذا سيدفع قمم المفردات، (إذا ما تخيلنا وجود المنحنى التكراري بصورة عمودية)، للاتجاه نحو اليسار مما يؤدي لاستطالة ذيل المنحنى من اليمين.

- يتم ترتيب قيم البيانات تصاعديا، ثم توضع هذه القيم بالترتيب (من الأصغر إلى الأكبر) في عامود الساق بعد حذف الخانة الأولى من يمين العدد. ويراعى عدم كتابة الأعداد المتشابهة بعد الحذف، وكتابة الأعداد الناقصة لاستكمال تسلسل الأعداد.
- يتم كتابة أعداد الخانة الأولى (التي تم حذفها من يمين العدد سابقا) في خانة الورقة بحيث ترتب تصاعديا من اليسار إلى اليمين.

ويمكن باستخدام شكل الساق والورقة حساب الوسيط والربيعات، وكذلك حساب القيم المتطرفة. وسنقوم بإيضاح طريقة تكوين شكل الساق والورقة بصورة عملية من خلال المثال التالي.

مثال (18.3): البيانات التالية تمثل عدد الساعات التي قضاها 21 شخصا على شبكة الانترنت خلال أحد الأشهر. والمطلوب استخدام شكل الساق والورقة لوصف البيانات.

25	31	17	15	40	32	27
13	9	14	21	20	19	20
7	45	36	39	60	11	10

الحل:

نقوم أولا بترتيب الأعداد تصاعديا:

07	09	10	11	13	14	15	17	19	20	20	21	25	27	31	32	36
39	40	45	60													

ثم نقوم بإنشاء عمودي الساق والورقة، (كما هو موضح في شكل (10.3))، ونرتب الأعداد تصاعديا في عامود الساق بعد حذف الخانة الأولى إلى اليمين بدون تكرار الأعداد.

الساق	الورقة						
0	7	9					
1	0	1	3	4	5	7	9
2	0	0	1	5	7		
3	1	2	6	9			
4	0	5					
5							
6	0						

شكل (10.3): شكل الساق والورقة لعدد الساعات المستغرقة على الانترنت، مثال (18.3).

نقوم بعد ذلك بكتابة الخانة الأولى في عامود الورقة بحيث نبدأ بالقيمة الصغرى إلى اليسار ونرتب الباقي في نفس الصف تصاعديا؛ فمثلا نضع القيمة 0 في خانة الساق لأنها الأقل ولا نكررها مرة أخرى رغم وجودها في

العدد 09، وهكذا حتى أكبر قيمة وهي 6 (المناظرة لـ 60)، ثم نقوم بكتابة الخانة الأولى في القيم الأصلية في عامود الورقة بحيث تكون القيمة مرتبة تصاعديا في نفس الصف، فمثلا في الصف الأول المناظر للقيمة 0 نكتب أولا 7 ثم 9 كما هو موجود في الشكل، وهكذا الحال بالنسبة لباقي الصفوف. ولاحظ أن القيمة 5 في عامود الساق لم يكتب أمامها أي قيمة لأنها غير موجودة أصلا في قيم المفردات.

ويمكن من شكل (10.3) ملاحظة وجود قيمة متطرفة، (وهي 60)، ضمن قيم البيانات، ويلاحظ أيضا وجود التواء بسيط موجب في توزيع البيانات، (إذا ما تم رسم المنحنى المنقط على شكل الساق والورقة). وتتركز البيانات حول القيمة 20 تقريبا، مع وجود "فجوة" في قيم البيانات من القيمة 45 إلى القيمة 60.

ولحساب الوسيط من شكل الساق والورقة نتبع الآتي؛ حيث أن عدد المفردات هو 21 وهو عدد فردي تكون رتبة الوسيط هي الحادية عشر، فنقوم بعد الأرقام في عامود الورقة ابتداء من أول قيمة وهي 7 حتى نصل للقيمة الحادية عشر وهي 0، (القيمة داخل المستطيل)، فتكون قيمة الوسيط هي حاصل دمج القيمة المناظرة في عامود الساق وهي 2 مع تلك القيمة، فيكون $\tilde{X} = 20$.

لحساب الربيع الأول، وهي القيمة التي تقع في منتصف النصف الأول من البيانات، (الإحدى عشر قيمة الأولى)، إضافة لقيمة الوسيط، فنبحث عن القيمة السادسة من البداية، (القيمة داخل الدائرة)، وهي القيمة 4 فتكون قيمة الربيع الأول هي ناتج دمج القيمة المناظرة في عامود الساق وهي 1 مع تلك القيمة فيكون $Q_1 =$ 14.

وبالمثل لحساب الربيع الثالث، وهي تلك القيمة التي تقع في منتصف النصف من البيانات، (الإحدى عشر قيمة الثانية)، إضافة لقيمة الوسيط، فنبحث عن القيمة السادسة ابتداء من قيمة الوسيط، (وهي القيمة داخل المعين)، وهي 2 فتكون قيمة الربيع الثالث $Q_3 = 32$. وهكذا فإن ملخص الأرقام الخمسة يكون:

$min(X)$	Q_1	$\tilde{X}$	Q_3	$max(X)$
7	14	20	32	60

وهذا يعني أن معدل عدد الساعات التي يقضيها هؤلاء الأشخاص على الانترنت هو 20 ساعة تقريبا، حيث أن الوسط الحسابي لعدد الساعات هو $\bar{X} = 24.33$ والذي من الواضح أنه تأثر بالقيمة المتطرفة 60. ومعظم هؤلاء الأشخاص، (75% منهم)، يقضون عدد ساعات لا يزيد عن 32 في الشهر الواحد، أي بمعدل ساعة واحدة في اليوم تقريبا.

ونختتم هذا الفصل بالإشارة إلى أنه لا توجد طريقة واحدة أو أسلوب إحصائي محدد يمكن استخدامه للتعامل مع كل أنواع البيانات، أو حتى نفس النوع من البيانات عندما تكون طبيعة المتغيرات مختلفة. كما أنه ليس من الحكمة الاعتماد "كليا" على بعض مقاييس النزعة المركزية والتشتت دون الاستعانة بالتمثيل البياني لأنه مثلا عند مقارنة المتغيرات قد تكون المقاييس المحسوبة أحيانا متقاربة، إلا أن هذا قد لا يعني بالضرورة أن توزيع هذه المتغيرات متشابه أو متقارب كما يوضح المثال القادم.

مثال[1] (19.3): قارن بين توزيعي المتغيرين التاليين A و B، المعطاة قيمهما كالتالي:

المتغير A							
0.40	0.40	0.40	0.41	0.42	0.42	0.43	0.44
0.44	0.44	0.47	0.49	1.98	1.99	1.99	2.00
2.10	2.11	2.14	2.17	2.20	2.23	3.00	3.10
3.20	3.23	3.30	3.40	4.90	5.00	5.00	5.10
6.60	6.70	6.77	6.80	6.90	7.00	7.77	7.80
7.83	7.86	7.89	7.90	8.00	8.01	8.01	8.02
9.51	9.53	9.56	9.56	9.56	9.57	9.58	9.58
9.59	9.60	9.60	9.60				

المتغير B							
−13.30	2.80	3.09	3.56	3.78	3.98	4.00	4.02
4.05	4.10	4.12	4.30	4.32	4.40	4.45	4.60
4.60	4.67	4.67	4.70	4.80	4.88	4.88	4.89
4.90	4.90	4.95	4.98	5.00	5.00	5.00	5.00
5.02	5.05	5.10	5.10	5.11	5.12	5.12	5.20
5.30	5.33	5.33	5.40	5.40	5.55	5.60	5.68
5.70	5.88	5.90	5.95	5.98	6.00	6.02	6.22
6.44	6.91	7.20	23.30				

الحل:

إذا ما تم حساب الوسط الحسابي والوسيط للمتغيرين فإن النتيجة ستكون

$$\bar{X}_A = 5,\ \bar{X}_B = 5.31 \text{ ، و } \tilde{X}_A = \tilde{X}_B = 5$$

والانحراف المعياري لكل منهما

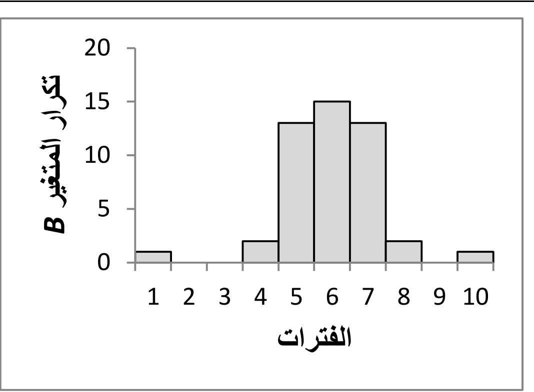
$$S.D_A = 3.44 \text{ ، و } S.D_B = 2.50$$

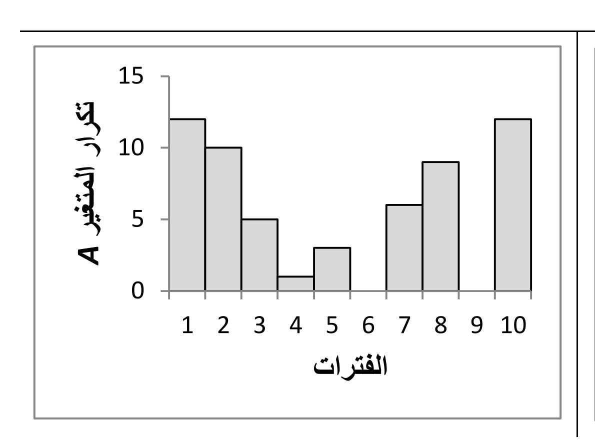
وهذا قد يدل على أن المتغيرين A و B لهما توزيع قيم متشابه، إلا أن الرسم البياني (المدرج التكراري[2] في الشكل (11.3)) لكل من المتغيرين يعكس صورة مختلفة تماما.

---

[1] البيانات من فرانك وألثوين (Frank and Althoen)، (1994).

[2] تم استخدام رموز متسلسلة (من 1 إلى 10) للفترات في كلا المدرجين لجعل المقارنة بين التوزيعين أسهل.





شكل (11.3): المدرج التكراري لبيانات المتغيرين A و B ، (مثال (19.3)).

فمن الشكل (11.3) يُلاحظ أن توزيعي المتغيرين مختلفين لدرجة أن أحدهما يكاد يكون عكس الآخر، فتوزيع البيانات في المتغير A له قمتين، والقيم الوسطى حول المركز (الفترات من 4 إلى 6) لها أقل تكرارات مناظرة، بينما تتمتع القيم الوسطى في المتغير B بأعلى تكرارات وشكل التوزيع ككل يكون أكثر تدببا و به قيمتين متطرفتين هما (13−) و (23).

وهذا يعني أن المقارنة بين توزيع المتغيرين A و B ما كانت لتعتبر صحيحة إذا ما تم الاكتفاء بالمقاييس المحسوبة من وسط وانحراف معياري دون الاستعانة بالتمثيل البياني.

4.3 تمارين الفصل الثالث

تمرين (1.3): للمفردات التالية، (جدول 1)، والتي تمثل معدلات درجات الحرارة في مدينة بنغازي في فصل الربيع مقاسة بالدرجة المئوية في إحدى السنوات:

جدول 1:

21	25	24	22	22	18	20	24	19	20

أ. أوجد:

1. المدى 2. المدى الربيعي 3. نصف المدى الربيعي 4. الانحراف المتوسط 5. الانحراف المعياري والتباين.

ب. مستخدما البيانات في جدول 1 و جدول 2 (والتي تمثل معدلات درجات الحرارة في مدينة الاسكندرية في نفس الفصل مقاسة بالفهرنهايت)، قارن بين معدلات درجات الحرارة في المدينتين باستخدام مقياس معامل الاختلاف.

جدول 2:

69.8	77	75.2	71.6	71.6	64.4	68	75.2	66.2	68

تمرين (2.3): الجدول التالي يمثل التوزيع التكراري لأعمار 300 شخصا في عدة سجون، (أحداث وبالغون)، يقضون مدة سجنهم بتهمة السرقة. أوجد:

1. المدى 2. نصف المدى الربيعي 3. الانحراف المتوسط 4. الانحراف المعياري والتباين. 5. المدى المئيني (10–90).

الفترة (الفئة العمرية)	5 – 15	15 – 25	25 – 35	35 – 45	45 – 55
التكرار f	25	120	105	31	19

تمرين (3.3): أوجد القيم المعيارية لدرجات الحرارة في جدول 1 وجدول 2 في التمرين (1.3) أعلاه مع التعليق. ثم احسب الوسط الحسابي والانحراف المعياري لكل منهما.

تمرين (4.3): إذا كان معدل الأجور الشهرية في الشركة A هو 900 دينار بانحراف معياري 320 دينار، وكان معدل الأجور في الشركة B هو 600 دينار بانحراف معياري 210 دينار، وتم اختيار موظف من الشركة A وآخر من الشركة B فكان راتب الأول هو 1000 دينار وراتب الثاني هو 850 دينار. فأي من الموظفين راتبه أفضل مقارنة بالشركة التي يعمل بها؟

تمرين (5.3): للبيانات التالية:

2	0	-6	12	7	9	-16	7	21	3

أوجد:

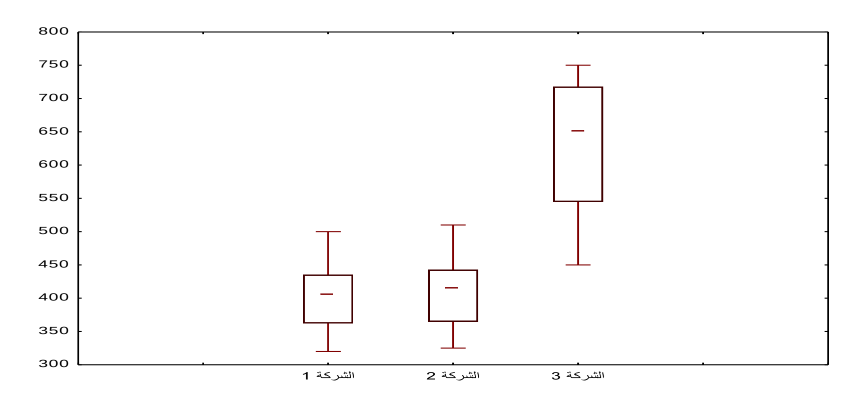
أ. العزوم الأربعة الأولى حول الصفر ب. الوسط الحسابي والتباين باستخدام العزوم ج. العزوم الأربعة الأولى حول الوسط الحسابي.

تمرين (6.3): البيانات التالية تمثل توزيع أعمار المصلين الذين يرتادون أحد مساجد مدينة اسطنبول في فترة معينة:

الفترة	5 - 15	15 - 25	25 - 35	35 - 45	45 -55	55 - 65
التكرار f	12	14	15	20	43	57

والمطلوب استخدام مقاييس الالتواء والتفرطح لوصف توزيع البيانات، (استخدم رسم المدرج التكراري لمزيد من التوضيح).

تمرين (7.3): الرسم البياني التالي يمثل شكل الصندوق لتوزيع أعداد المشتركين في 3 شركات لخدمات الانترنت خلال 12 شهرا في إحدى المدن خلال عام 2013. ما هو التعليق المناسب؟.



تمرين (8.3): استخدم شكل الساق والورقة لوصف البيانات التالية والتي تمثل درجات ذكاء 25 طفل في المرحلة الابتدائية، علما بأن المقياس هو من 100 درجة.

20	30	33	40	44	45	50	52	53
55	60	61	65	62	65	66	70	72
74	78	84	88	89	91	92		

الفصل الرابع

أساسيات الاحتمال

(Fundamentals of Probability)

1.4 مقدمة (Introduction)

2.4 الأحداث وفراغ العينة (Events and Sample Space)

1.2.4 التجربة العشوائية والحدث (Random Experiment and Event)

2.2.4 فراغ العينة ونظرية الفئات (Sample Space and Set Theory)

3.2.4 بعض العمليات الأساسية على الفئات (Some Basic Operations on Sets)

3.4 طرق العد ومسلمات الاحتمال (Counting Methods and Probability Axioms)

1.3.4 طرق العد (Counting Methods)

2.3.4 مسلمات الاحتمال (Probability Axioms)

3.3.4 النظريات الأساسية للاحتمال (Basic Theorems for Probability)

4.4 الاحتمال الشرطي ونظرية بيز (Conditional Probability and Bayes Theorem)

4.5 تمارين الفصل الرابع

1.4 مقدمة (Introduction)

تناولنا في الفصل الأول الطرق المختلفة لاختيار العينات من المجتمع، ووضحنا أن التعامل مع بيانات العينة يكون ضروريا في أغلب الأوقات نظرا لصعوبة التعامل مع بيانات المجتمع ككل بسبب ضخامة المجتمع و/أو ارتفاع تكاليف جمع بيانات المجتمع كلها، وغيرها من الأسباب. وأوضحنا أيضا ضرورة تمتع العينة المسحوبة بخاصية العشوائية والتي تضفي على مفردات العينة مصداقية أكبر لتمثيل مفردات المجتمع، وهذه الطبيعة العشوائية هي في الواقع السبب الأساسي في التغيّر في البيانات، ويعتبر المحفز الرئيسي لاستخدام علم الإحصاء وأدواته.

وعلم الاحتمالات يهدف لدراسة هذا التغير في أساليب اختيار العينات وتوظيفه في مجال الاستدلال الإحصائي، بحيث يحدد لنا الدرجة التي قد نصل إليها في ثقتنا بهذا الاستدلال عن مفردات المجتمع من خلال العينة المسحوبة.

في حياتنا اليومية، كثيرا ما نستخدم في حديثنا عبارات لا تحمل في مضمونها تأكيدا صريحا أو محددا، عبارات مثل "أظن بأني سوف أتحصل على تقدير 75% في مادة الإحصاء هذا العام"، "هذا التلفاز من المفترض أن يعمل 5 سنوات بدون أعطال"، و "من المتوقع أن يكون الطقس غائما غدا مع احتمال هطول أمطار خفيفة ليلا"، ...، وهكذا. مثل هذه العبارات يمكن أن تصنف بأنها حدس استقرائي (Intuitive Induction) أو أحكام مسبقة يصدرها الناس بناء على توقعات معينة مرتبطة أو غير مرتبطة بتجارب أو خبرات سابقة وليست حساب الاحتمالات.

إن ما يفرق المفهوم العام لمصطلح "احتمال (Probability)" عن مفهوم "الاحتمال الإحصائي (Statistical Probability)" هو أن الأخير لا يستند إلى الحدس أو الحكم الشخصي على الأشياء، بل يعتمد على معرفة كل النتائج الممكنة للظاهرة أو التجربة. وعند التعامل مع النتائج أو المشاهدات المحيطة بنا فإننا إحصائيا نكون مهتمين بفهم هذه المشاهدات ومحاولة التوصل لطريقة رياضية لقياسها، ومن ثمة القيام باستخدامها في التنبؤ أو الاستدلال مستقبليا.

من ناحية أخرى، فإن تلك المشاهدات قد لا يمكن أحيانا التنبؤ بها أو توقع حدوثها بالتأكيد عند نقطة معينة، فمثلا نسبة السكر في الدم لشخص ما عند نقطة زمنية محددة لا يمكن توقعها بصورة "مؤكدة"، وكذلك من الصعب معرفة مقدار قوة تحمل نوع جديد من الحبال "بالضبط" قبل أن ينقطع أثناء التجربة. إن مثل هذه الظواهر أو التجارب لا يمكن التنبؤ بمفرداتها بشكل مؤكد عند نقطة محددة، ولكن من الملاحظ أن تكرار وقوع أو حدوث هذه النتائج خلال سلسلة طويلة نسبيا من المحاولات غالبا ما يكون مستقرا، والمشاهدات الناتجة عن هذه المحاولات تسمى "**أحداثا**"[1] عشوائية أو تصادفية (Stochastic).

وهنا يأتي دور النظرية الرياضية في توظيف هذا التكرار الناتج من المحاولات ومحاولة صياغته في شكل أو إطار يتيح للباحث حساب احتمال حدوث نتيجة معينة، فمثلا قد يكون من المستحيل التأكد من الحصول على صورة (Head) عند رمي قطعة نقود (عملة) معدنية لمرة واحدة، ولكن يمكن حساب احتمال الحصول على

[1] سنأتي لتعريف مصطلح "**الحدث**" في علم الاحتمالات لاحقا في هذا الفصل.

الصورة خلال سلسلة من محاولات رمي تلك العملة[1]. ونظرية الاحتمال هي أحد فروع علم الرياضيات التطبيقي التي تهتم بدراسة تأثير الصدفة (Chance) على الأشياء من حولنا.

إن مفهوم الاحتمال القديم، والذي ظهر خلال عدة حضارات سابقة، يعود لاهتمام الناس الدائم بالحظ والصدفة والفأل الحسن والسيئ وغيرها من الموروثات الاجتماعية. فقد ذكر المؤرخون أن قدماء المصريين استخدموا قطعة من العظم لها أربعة أوجه كزهر نرد بدائي قبل ظهوره بأوجهه الستة في العام 1600 قبل الميلاد. وفي عهد الإسلام وردت في القرآن الكريم آيات توضح صورا كانت سائدة في المجتمع الجاهلي تتعلق بالقمار وطلب الفأل والاعتماد على الحظ في اتخاذ القرارات[2]. وعموما، فإن ظهور الاحتمالات كعلم رياضي كان في القرن الخامس عشر على يد علماء مثل "باتشولي (Paccioli)" و "جاليلي (Galilei)" والذين قاموا بحساب الاحتمالات المتعلقة بألعاب الحظ المختلفة.

أما التطور الحقيقي فقد بدأ في فرنسا في العام 1654 عن طريق "باسكال (Pascal)" و "فيرمات (Fermat)"، مرورا "ببيرنوللي (Bernoulli)" في العام 1713 و "لا بلاس (La Place)" و "بواسون (Poisson)" و "جاوس (Gauss)" بعد ذلك، وتتابعت تطورات علم الاحتمالات في القرن التاسع عشر من خلال العالمين "تشيبيتشف (Chebyshev)" و "ماركوف (Markov)". ومنذ أوائل القرن العشرين أصبحت الاحتمالات مجموعة من النظريات الرياضية الحديثة القائمة على أساس إحصائي.

وبمساعدة نظريات وقوانين علم الاحتمالات ازدادت قدرتنا على الإسهام في المجالات العلمية والتقنية الحديثة مثل الطب والبيولوجيا، علم المعلوماتية (Information Technology)، الملاحة الفضائية، والتسويق التجاري، إضافة إلى استخدامات الاحتمالات في مجالات السياسة، الاقتصاد والأعمال، الأرصاد الجوية وغيرها.

أما عن أهمية علم الاحتمالات الغير مباشرة في تطوير علم الإحصاء التقليدي فإنه ببساطة شديدة يمكن اعتبار الاحتمالات "كحلقة وصل" بين قطبي علم الإحصاء الأساسيين؛ الاستكشافي والاستدلالي، وأن نظرية الاحتمال أسهمت بشكل جذري في وضع أسس علم الإحصاء الاستدلالي الذي يعتمد على التوزيعات الاحتمالية وخواصها كما سنرى في الفصول اللاحقة.

2.4 الأحداث وفراغ العينة (Events and Sample Space)

في الجزء السابق تم تسليط الضوء على نشأة علم الاحتمالات ومراحل تطوره، وناقشنا باختصار أهمية علم الاحتمالات بالنسبة لعلوم الحياة الأخرى وأهم تطبيقاته الحديثة. وفي هذا الجزء سنقوم بتعريف بعض المفاهيم والخواص الرئيسية التي تشكل الأساس الرياضي الذي يبنى عليه هذا العلم. وتعتبر الأحداث وفراغ العينة من أهم تلك المفاهيم.

---

[1] والذي سيكون حول القيمة $\frac{1}{2}$.

[2] نذكر منها "يا أيها الذين آمنوا إنما الخمر والميسر والأنصاب والأزلام رجس من عمل الشيطان فاجتنبوه لعلكم تفلحون"، المائدة (90). حيث أن الميسر هو لعب القمار، والأنصاب والأزلام هي طرق لعمل الاقتراع والاستقسام وطلب الفأل.

1.2.4 التجربة العشوائية والحدث (Random Experiment and Event)

إن مصطلح "تجربة" يعد من المصطلحات الكثيرة الاستخدام في عالمنا، ولكي نكون أكثر دقة فإنه يجب التفريق بين مفهومي التجربة والتجربة العشوائية، فالأول قد يمثل محاولة يتم من خلالها إجراء عملية ما تحت ظروف قياسية ثابتة يتم السيطرة عليها من قبل الباحث لضمان الحصول على نتيجة نهائية "مؤكدة"، كعملية تركيب دواء كيميائيا عن طريق خلط عدة مقادير بنسب ثابتة محددة سلفا.

أما مصطلح "التجربة العشوائية" فهو يستخدم عادة في علم الإحصاء لوصف أي عملية أو ظاهرة نتيجتها المباشرة غير محددة أو معلومة بالتحديد مسبقا، فتحديد عدد خلايا البكتيريا الموجودة مثلا في السنتيمتر المكعب على قطعة من الخبز، واختيار طالب (بدون تمييز) من فصل ما لإلقاء خطاب نيابة عن الطلبة في حفل سنوي، ومعرفة من سيفوز في مباراة لكرة القدم، كلها تعرف بأنها تجارب عشوائية، لأن النتيجة الفعلية لا يمكن تحديدها بصورة مؤكدة أو قاطعة مسبقا. ونسوق فيما يلي التعريف الإحصائي لمفهومي التجربة والتجربة العشوائية:

تعريف (1.4): **التجربة** (Experiment): التجربة هي العملية التي تُنفذ للكشف عن حقيقة معينة، ويتم عن طريقها تحديد المشاهدات الناتجة (البيانات).

تعريف (2.4): **التجربة العشوائية**[1] (Random Experiment): التجربة العشوائية هي التجربة التي تولد مجموعة من المشاهدات بحيث لا يمكن الجزم بالنتيجة المحددة التي ستؤول إليها هذه التجربة من ضمن مجموعة النتائج الكلية الممكن الحصول عليها.

عند إجراء أي تجربة عشوائية، فإننا نحصل عادة على مجموعة من المشاهدات والتي يمكن أن نطلق عليها تسمية "أحداث" في علم الاحتمالات، فمثلا؛ في تجربة رمي زهر نرد لمرة واحدة يمكننا الحصول على عدة أحداث يمكن تمثيلها كالتالي:

A يمثل حدث الحصول على الرقم 5، B يمثل حدث الحصول على رقم فردي، و C يمثل حدث الحصول على رقم أكبر من 2. هذه الأحداث يمكن التعبير عنها بصورة رياضية أكثر بساطة بالصورة التالية: $A = \{5\}$ ، $B = \{1, 3, 5\}$ ، و $C = \{3, 4, 5, 6\}$ ، وهكذا.

ونلاحظ من هذه التجربة أن بعض الأحداث، وهي B و C، تحتوي على أكثر من مشاهدة (عنصر) ضمن نتائجها، وأن الحدث A يحتوي على عنصر واحد فقط، في هذه الحالة يسمى الحدث A بالحدث البسيط، وأما الأحداث B و C فتسمى أحداثا مركبة. ويمكن تعريف الحدث بالصورة التالية:

تعريف (3.4): **الحدث** (Event): الحدث هو النتيجة التي يتم الحصول عليها عند إجراء أي تجربة عشوائية.

---

[1] بعض كتب علم الاحتمالات تستخدم مصطلح "التجربة الاحتمالية (Probability Experiment)" إضافة لمصطلح التجربة العشوائية.

فإذا كانت هذه النتيجة تحتوي على عنصر واحد لا يمكن تقسيمه إلى مجموعات أخرى سمي هذا الحدث **بالحدث البسيط** (Simple Event)، أما إذا أمكن تقسيم نتيجة التجربة العشوائية إلى مجموعات أخرى فإن الحدث يسمى **بالحدث المركب** (Compound Event).

وفيما يلي سنتناول أهم العمليات الرياضية التي تنظم العلاقات بين الأحداث الناتجة عن التجارب العشوائية المختلفة، وذلك بهدف تسهيل حساب الاحتمالات بصورة رياضية ومنطقية.

2.2.4 فراغ العينة ونظرية الفئات (Sample Space and Set Theory)

في تعريفنا للتجربة العشوائية، ذكرنا أنه لا يمكن تحديد أي نتيجة معينة ستؤول إليها التجربة مسبقا، إلا أننا نعلم مسبقا جميع النتائج التي ستظهر من خلال هذه التجربة؛ فمثلا عند رمي قطعة نقود معدنية متزنة[1] مرة واحدة فإننا نعلم جميع نتائج هذه التجربة وهي الصورة (Head) والكتابة (Tail)[2]، ولكننا لا نعلم بالتحديد ما سنحصل عليه بالضبط، الصورة أم الكتابة.

لذلك فإنه من المفيد جدا لدراسة هذه التجارب العشوائية تحديد المجموعة أو الفئة (Set) التي تضم كل النتائج المتوقعة للتجربة، ومن ثمة تعريف الأحداث عليها. فإذا ما اعتبرنا أن الحدث هو مجموعة تضم نقطة أو مجموعة من نقاط العينة (Sample Points) المسحوبة من جميع النتائج الممكنة للتجربة، فإنه يمكن تعريف فراغ العينة، والذي يرمز له بالرمز S، كالتالي:

تعريف (4.4): **فراغ العينة** (Sample Space): فراغ العينة لأي تجربة عشوائية هو المجموعة أو الفئة التي تحتوي على كل النتائج الممكنة (نقاط العينة) لتلك التجربة، ويرمز لعدد العناصر في فراغ العينة بالرمز $n(S)$.

ونسرد بعض الأمثلة التوضيحية لفراغ العينة فيما يلي:

مثال (1.4):

- عند رمي زهر نرد مرة واحدة، فإن فراغ العينة لكل النتائج الممكنة يكون؛
  $S = \{1, 2, 3, 4, 5, 6\}$ ، وعدد عناصر S هو $n(S) = 6$.
- يمكن إلقاء عملة معدنية وزهر نرد في تجربة واحدة فيكون فراغ العينة للنتائج؛
  $S = \{H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6\}$ ، ويكون $n(S) = 12$.
- لعائلة رزقت بطفلان يمكن تكوين فراغ العينة لجنس المولود باعتبار الطفل الأكبر أولا ثم الأصغر ثانيا كما يلي:

---

[1] يقصد بها أن قطعة النقود سليمة وغير مغشوشة، (كبعض القطع التي يستخدمها بعض لاعبي القمار في خداع الآخرين)، وأن كل وجه من وجهي القطعة المختلفين له نفس فرصة الظهور.

[2] عادة ما يستخدم الحرف H للدلالة على الصورة، والحرف T للدلالة على الكتابة.

$S = \{BB, BG, GB, GG\}$ ، حيث يرمز B للطفل الذكر و G لأنثى، فإما أن يكون الطفلين كلاهما من الذكور (BB)، أو كلاهما من الإناث (GG)، أو يكون الأول ذكرا والثانية بنتا (BG) أو العكس (GB). ويكون $n(S) = 4$.

وبصورة عامة، إذا ما رمزنا لأي حدث ضمن تجربة عشوائية بالرمز E_i ، بحيث يحتوي فراغ العينة على n نتيجة كلية (عنصر)، $i = 1, 2, ..., n$ ، عندئذ نستطيع كتابة: $S = \{E_1, E_2, ..., E_n\}$ ، حيث $n(S) = n$.

في كثير من الأحيان، قد نصادف بعض التجارب المركبة أو المعقدة والتي تم إجراؤها عبر سلسلة من المراحل، هذه التجارب قد يكون من الصعوبة تكوين فراغ العينة لها بطريقة مباشرة، لذلك فإننا نلجأ لاستخدام ما يعرف **بمخطط الشجرة البيانية** (Tree Diagram) لتوضيح مخطط سير التجربة العشوائية، وبالتالي حصر جميع النتائج الممكنة لها.

مثال (2.4): في جامعة بنغازي، تم اختيار ثلاثة طلبة من إحدى الكليات بغرض إجراء دراسة ميدانية اجتماعية حول معدلهم الدراسي، (دراسة حالة). كون فراغ العينة لتجربة اختيار الطلبة من حيث كون الطالب المختار (عشوائيا) ذكرا أو أنثى.

الحل:

لنفرض أن B = حدث أن الطالب المختار ذكر، و G = حدث أن الطالب المختار أنثى، فيكون لدينا الخيارات التالية لاختيار الطلبة الثلاثة:

اختيار الطالب الثالث	اختيار الطالب الثاني	اختيار الطالب الأول
B	*B*	*B*
G		
B	*G*	
G		
B	*B*	*G*
G		
B	*G*	
G		

ويتم تكوين فراغ العينة من خلال تتبع اختيار المشاهدات عبر فروع الشجرة البيانية فنحصل على $n(S) = 8$ عناصر هي: $S = \{BBB, BBG, GB, BGB, BGG, GBB, GBG, GGB, GGG\}$.

ولاحظ أن العدد الكلي لعناصر فراغ العينة هو عبارة عن $8 = (2)^{(3)}$ حالات كلية، بمعنى أنه يمكن، بصورة عامة، كتابة:

عدد عناصر فراغ العينة = (عدد الحالات في المحاولة الواحدة)$^{(\text{عدد مرات تكرار المحاولات})}$

3.2.4 بعض العمليات الأساسية على الفئات (Some Basic Operations on Sets)

تعريف (5.4): **الفئة الخالية** (Empty Set): الفئة الخالية هي تلك الفئة التي لا تحتوي على أي عنصر داخلها، ويرمز لها بالرمز Φ، أي أن $n(S) = 0$.

من خلال هذا التعريف (5.4) نستطيع تعريف **الفراغ الخالي** (Empty Space) أو **فراغ العدم** (Null Space) بأنه فئة جزئية من أي فراغ عينة، وهو لا يحتوي على أي حدث، بمعنى أن $\Phi \subset S$.

تعريف (6.4): **الاتحاد** (Union): إذا كان A و B هما أي حدثين معرفين على فراغ العينة S، فإن اتحاد A وB يعرف بأنه الحدث الذي يحتوي على كل النتائج في A أو في B أو في كلاهما، ويرمز لتلك العملية بالرمز $A \cup B$، ونستطيع كتابة:

$$A \cup B = \{x: x \in A \text{ أو } x \in B\}$$

وتتمتع عملية الاتحاد بين الأحداث بالخواص التالية:

لنفرض أن $A, B \subset S$ عندئذ يكون

$$A \cup \Phi = A \;\;,\;\; A \cup A = A \;\;,\;\; A \cup S = S \;\;,\;\; A \cup B = B \cup A$$

وإذا كان $A \subset B$ فإن $A \cup B = B$.

والاتحاد بين الفئات A، B، و C حيث $A, B, C \subset S$ يمكن وضعه بالصورة:

$$A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C)$$

وهذه الخاصية الأخيرة يمكن تعميمها كالتالي؛

إذا كانت $A_1, A_2, \dots, A_k$ هي أحداث معرفة على فراغ العينة S فإن

$$A_1 \cup A_2 \cup \dots \cup A_k = \bigcup_{i=1}^{k} A_i$$

تعريف (7.4): **التقاطع** (Intersection): إذا كان $A, B \subset S$ فإن تقاطعهما يعرف بأنه الحدث الذي يحتوي كل النتائج في A و B. ويرمز لتلك العملية بالرمز $A \cap B$، ويعبر عن ذلك بالصورة؛

$$A \cap B = \{x: x \in A \text{ و } x \in B\}$$

ومن خواص عملية التقاطع أن:

$$A \cap \Phi = \Phi \;\;,\;\; A \cap A = A \;\;,\;\; A \cap S = A \;\;,\;\; A \cap B = B \cap A$$

وإذا كان $A \subset B$ فإن $A \cap B = A$. وعموما، إذا كانت $A_1, A_2, \dots, A_k$ هي أحداث معرفة على فراغ العينة S فإن تقاطعها يعرف بالصورة:

$$A_1 \cap A_2 \cap \dots \cap A_k = \bigcap_{i=1}^{k} A_i$$

تعريف (8.4): **الحدث المكمل** (Complement Event): يعرف مكمل الحدث A بأنه الحدث الذي يحتوي على كل العناصر الموجودة في فراغ العينة S وغير موجودة في الحدث A، ويرمز له بالرمز A^c، ويكتب بالصورة الرياضية التالية:

$$A^c = \{x: x \in S \text{ و } x \notin A\}$$

ويكون

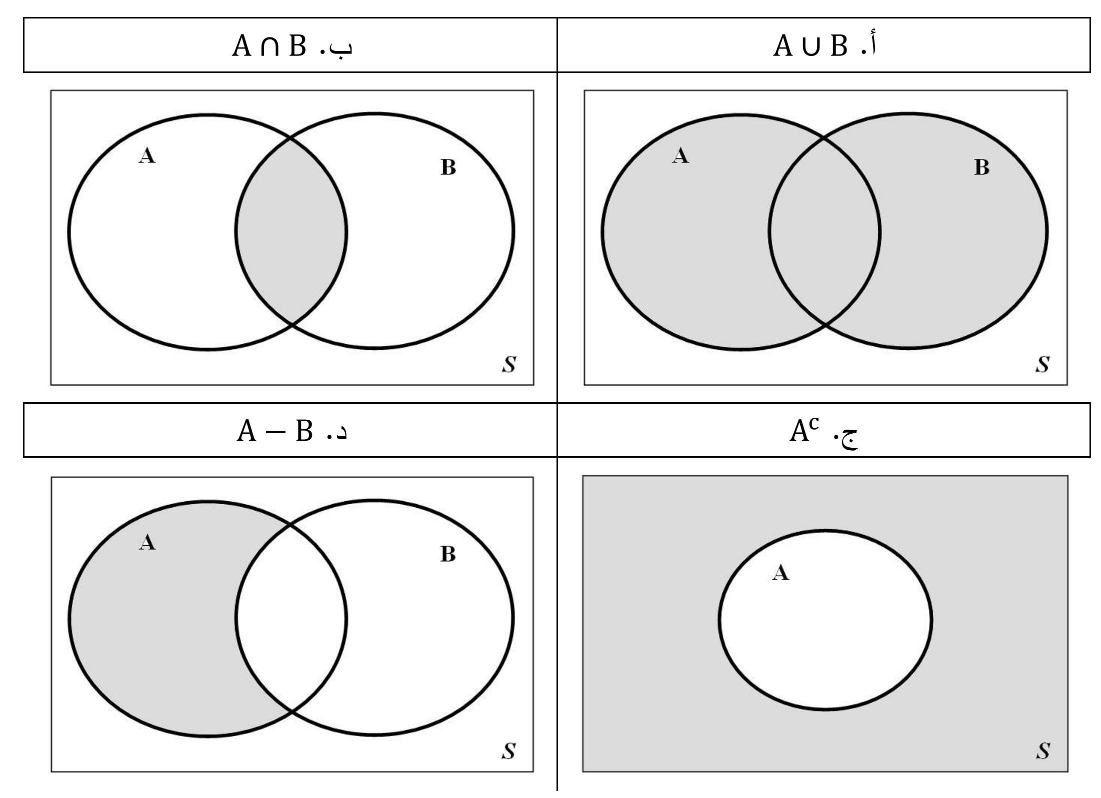
$$(A^c)^c = A \;, A \cup A^c = S \;, A \cap A^c = \Phi \;, \Phi^c = S \;, S^c = \Phi$$

تعريف (9.4): **الأحداث المتساوية** (Equal Events): نقول أن الحدثين A و B متساويين إذا كان $A \subset B$ و $B \subset A$ ، ويرمز للأحداث المتساوية بالرمز $A = B$.

تعريف (10.4): **الفرق بين الأحداث** (Difference between Events): الفرق بين الحدثين A وB، والذي يرمز له بالرمز $(A - B)$، هو ذلك الحدث الذي يحتوي على كل العناصر الموجودة في A وغير موجودة في B.

ونلاحظ أن $A - B = A \cap B^c$ و $A^c = S - A$.

والشكل (1.4) يوضح بيانيا، باستخدام مخططات أو **أشكال فن** (Venn Diagrams)، طبيعة العلاقات السابقة بين الأحداث.



شكل (1.4): أشكال فن لبعض العمليات الأساسية بين الأحداث.

تعريف (11.4): **القانون التوزيعي** (Distributive Law): إذا كان A، B، و C هي أحداث معرفة على فراغ العينة S فإن:

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

و $$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

تعريف (12.4): **قانون دي مورجان الأول** (De Morgan's First Law): إذا كان A و B أي حدثين في S فإن:

$$(A \cup B)^c = A^c \cap B^c$$

ويمكن تعميم العلاقة الأخيرة لتصبح بالصورة:

$$\left(\bigcup_{i=1}^{k} A_i\right)^c = \bigcap_{i=1}^{k} A_i^c \quad , \quad \left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c$$

تعريف (13.4): **قانون دي مورجان الثاني** (De Morgan's Second Law): لأي حدثين A و B في S يكون:

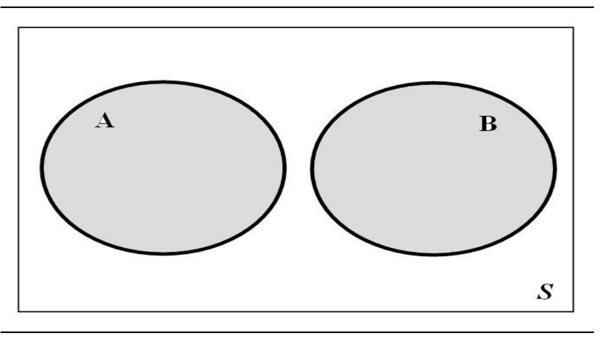
$$\left(\bigcap_{i=1}^{k} A_i\right)^c = \bigcup_{i=1}^{k} A_i^c \quad , \quad \left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c$$

وكنتيجة لبعض القوانين السابقة، يمكننا لأي حدثين A و B في S كتابة:

$$A = (A \cap B) \cup (A \cap B^c)$$

تعريف (14.4): **الأحداث المتنافية** (Mutually Exclusive or Disjoint Events): إذا كان A و B هما أي حدثين في فراغ العينة S، فإننا نقول أن A و B حدثان متنافيان إذا كان من غير الممكن وقوعهما في آن واحد، بمعنى إذا كان وقوع أحدهما ينفي وقوع الآخر.

فمثلا، في تجربة إلقاء عملة معدنية مرة واحدة فإن الحدث A = H والذي يمثل ظهور الصورة والحدث B = T والذي يمثل ظهور الكتابة، يكونا حدثين متنافيين إذ أنه لا يمكن الحصول على صورة وكتابة معا في آن واحد أو في رمية واحدة. والشكل (2.4) يوضح الرسم البياني لحدثين متنافيين.



شكل (2.4): شكل فن لحدثين متنافيين.

تعريف (15.4): **الأحداث المستقلة** (Independent Events): يقال أن الأحداث $A_1, A_2, ..., A_k$ هي أحداث مستقلة في فراغ العينة S إذا كان وقوع أحدها أو بعضها لا يؤثر على وقوع باقي الأحداث.

فعلى سبيل المثال، في تجربة إلقاء عملة معدنية مرتين[1] يكون فراغ العينة:

S = {HH, HT, TH, TT}. في هذه الحالة، إذا كان الحدث A_1 يمثل ظهور الصورة في العملة الأولى (أو الرمية الأولى)، و كان الحدث A_2 يمثل ظهور الصورة في العملة الثانية، فإن وقوع الحدث A_1 لا يمنع أو يؤثر على وقوع الحدث A_2 ، وبالتالي نقول أن A_1 و A_2 هما حدثين مستقلين.

مثال (3.4): لنفرض أنه تم تعريف فراغ عينة على أنه المجموعة التي تحتوي على كل أحرف اللغة العربية الثماني والعشرون، ($n(S) = 28$)؛ { أ، ب، ت، ث، ... ، ي } = S ، وتم تعريف الأحداث التالية على S :

A تمثل الأحرف المكونة لكلمة "أحمد"، B تمثل الأحرف المكونة لكلمة "خديجة"، و C تمثل الأحرف المكونة لكلمة "ريم". والشكل (3.4) يوضح أشكال فن لهذه الأحداث.

نلاحظ هنا أن عدد العناصر في كل فئة أو حدث هو $n(A) = 4$، $n(B) = 5$، و $n(C) = 3$. حيث { أ، ح، م، د } = A، { خ، د، ي، ج، ت[2] } = B، و { ر، ي، م } = C. عندئذ نستطيع، على سبيل المثال، تكوين العلاقات التالية بين الأحداث السابقة:

$A \cup B$ = { أ، ت، ج، ح، خ، د، م، ي }

$A \cap C$ = { م }

$A \cup B \cup C$ = { أ، ت، ج، ح، خ، د، ر، م، ي }

$A \cap B \cap C = \Phi$

$A \cap S = C$ = { ر، م، ي }

$A^c = S -$ { أ، ح، د، م }

$A - C$ = { أ، ح، د }

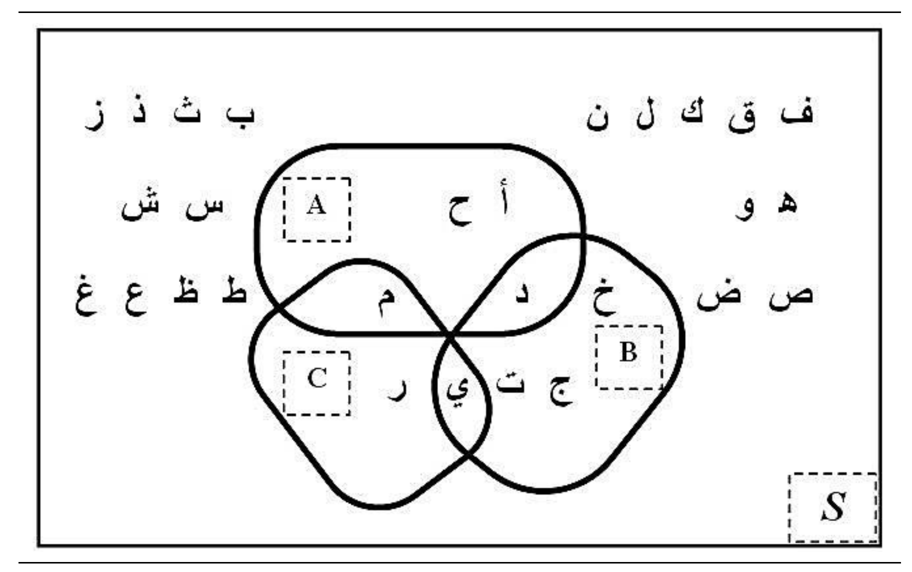
$(A \cup B)^c$ = { ب، ث، ذ، ر، ز، ... ، ل، ن، ه، و }

$(A \cup B)^c \cap C$ = { ر }

$(A \cup C)^c = A^c \cap B^c$ = { ب، ت، ث، ج، خ، ذ، ز، س، ش، ... ، ل، ن، ه، و }

[1] وهو ما يعادل، عمليا، إلقاء عملتين مرة واحدة.

[2] هنا نعتبر أن التاء المربوطة في كلمة "خديجة" تعادل التاء المفتوحة عند التعامل مع حروف الهجاء.



شكل (3.4): أشكال فن للأحداث A، B، و C ،(مثال (3.4)).

3.4 طرق العد ومسلمات الاحتمال (Counting Methods and Probability Axioms)

في البند السابق، تم مناقشة مفهوم التجربة العشوائية وكيفية تكوين فراغ العينة بناء على النتائج المتحصل عليها من تلك التجربة، ثم تم تعريف الحدث كفئة في فراغ العينة، وتناولنا أنواع الأحداث وأهم العلاقات الرياضية الأساسية التي تربط بينها. وفيما يلي، سنستعرض القواعد العامة التي تنظم تكوين فراغ العينة عند التعامل مع التجارب العشوائية الأكثر تعقيدا، والتي تعرف بطرق العد.

1.3.4 طرق العد (Counting Methods)

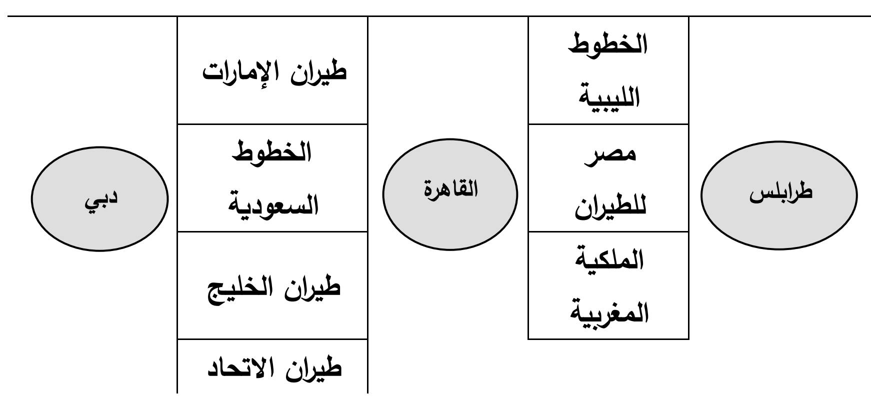
إن الاحتمال، كما سنرى لاحقا، لا يتم حسابه عادة بصورة مباشرة اعتمادا على سرعة البديهة، (كما هو الحال في بعض المسابقات)، بل يرتكز على قواعد وأسس رياضية تنظم وتضبط هذه الحسابات. هذه القواعد أو الطرق تسمى بطرق العد[1]، وهي إضافة لاستخداماتها في علم الإحصاء تعتبر من الأدوات الهامة في حقول أخرى مثل المعلوماتية، البرمجة الخطية، التخطيط الصناعي وغيرها.

وترتكز طرق العد على قاعدة رياضية أساسية تعرف **بقاعدة الضرب**، والتي سيتم تعريفها بعد المثال التوضيحي التالي، والذي نسوقه لإبراز أهمية استخدام طرق العد لتحليل التجارب العشوائية:

لنفرض أن أحد الأشخاص يرغب في السفر من مدينة طرابلس في ليبيا إلى مدينة دبي في الإمارات العربية المتحدة، على أن يقضي بضعة أيام في مدينة القاهرة بجمهورية مصر أثناء الرحلة (ترانزيت). وكان بإمكان هذا الشخص اختيار الحجز من بين ثلاث شركات طيران مختلفة للرحلة من طرابلس إلى القاهرة، والاختيار من بين أربعة شركات طيران أخرى للرحلة من القاهرة إلى دبي، (كما هو موضح في شكل (4.4)).

في هذه الحالة، فإن هذا المسافر لديه عدد طرق كلية للسفر تساوي حاصل ضرب كل الاختيارات المتاحة في بعضها البعض، وهي $3 \times 4 = 12$ طريقة كلية للوصول للوجهة المطلوبة، (مرورا بالقاهرة).

[1] قد يستخدم بعض الكتاب مصطلح **الطرق التجميعية أو التركيبية** (Combinatorial Methods).



شكل (4.4): المسارات المختلفة التي يمكن اختيارها للسفر من طرابلس لدبي.

التجربة السابقة، والتي تم تحليلها اعتمادا على المنطق فقط، يمكن التعامل معها باستخدام النظرية التالية:

نظرية (1.4): **قاعدة الضرب** (Multiplication Rule):

إذا تم إجراء تجربة ما بعدد n_1 طريقة، ثم تم إجراء تجربة ثانية بعدد n_2 طريقة، فإن التجربتين يمكن إجراؤهما معا بعدد $n_1 \times n_2$ طريقة.

وبتطبيق هذه النظرية على المثال التوضيحي السابق والخاص باختيار شركات الطيران، نجد أن المسافر سيكون لديه $n_1 \times n_2 = 3 \times 4 = 12$ طريقة كلية للسفر.

والنظرية (1.4) يمكن تعميمها لتصبح بالصورة التالية:

نظرية (2.4): **قاعدة الضرب العامة** (Generalized Multiplication Rule):

إذا تم إجراء تجربة ما بعدد n_1 طريقة، وتجربة ثانية بعدد n_2 طريقة، وتجربة ثالثة بعدد n_3 طريقة، ... ، وهكذا حتى إجراء التجربة رقم k بعدد n_k طريقة، فإن هذه التجارب كلها مجتمعة يمكن إجراؤها معا بعدد $n_1 \times n_2 \times \ldots \times n_k$ طريقة كلية.

مثال (4.4): بكم طريقة يمكن تأثيث منزل مكون من التقسيمات التالية؛ غرفة نوم، غرفة معيشة، غرفة استقبال، ومطبخ ؟. إذا علمت أنه يمكن اختيار الأثاث من معرض صغير للأثاث يحوي 3 غرف نوم، 4 غرف معيشة، 5 غرف استقبال، ومطبخين.

الحل:

باستخدام النظرية (2.4) يكون العدد الكلي للاختيارات الممكنة لتأثيث المنزل من ضمن الاختيارات المتاحة هو $n_1 \times n_2 \times n_3 \times n_4 = 3 \times 4 \times 5 \times 2 = 120$ طريقة.

نظرية (3.4): الفئة التي تحتوي على n عنصر، يمكن تقسيمها إلى 2^n فئة جزئية.

فعلى سبيل المثل، يمكن تقسيم الفئة (الحدث) $A = \{a, b, c\}$ إلى $2^3 = 8$ فئات أو أحداث جزئية هي $\{a\}$، $\{b\}$، $\{c\}$، $\{a, b\}$، $\{a, c\}$، $\{b, c\}$، $\{a, b, c\}$، و Φ [1].

مثال (5.4): يفخر أحد محلات بيع المثلجات (الأيس كريم) بأنه يقدم أكثر من 30 طبق مختلف من المثلجات، حيث من الممكن اختيار أي تركيبة من المكونات الخمس المختلفة للنكهات؛ الفانيليا، الشيكولاته، الفراولة، الموز، والليمون. فهل تؤيد قول صاحب هذا المحل؟.

الحل:

حيث أنه يمكن تكوين أي تركيبة من المكونات الخمس في المحل، فإن عدد الأطباق المختلفة يكون $2^5 = 32$ تركيبة (طبق). وحيث أن هذه التركيبات (الفئات الجزئية) تشمل الفئة الخالية، والتي تمثل طبقا فارغا في هذه الحالة، فإنه يمكن تكوين 31 طبقا مختلفا، وهكذا فإننا نؤيد قول صاحب المحل.

ملاحظة: يمكن تعميم القاعدة التالية لترتيب سحب العناصر:

يمكن اختيار r عنصر من n عنصر **بالإرجاع**[2] (With Replacement) و**اعتبار الترتيب** بعدد n^r طريقة، بحيث $r \leq n$. ويقصد بالإرجاع هنا أن كل عنصر يمكن اختياره ثم إرجاعه لفراغ العينة ثم الاختيار من جديد، وهو ما يعادل تكرار نفس العنصر خلال عملية السحب. أما اعتبار الترتيب فتعني إمكانية اختيار عنصرين أو أكثر بترتيبات مختلفة.

عند استخدام طرق العد، عادة ما نكون مهتمين بحساب العدد الكلي لمكونات أو عناصر فراغ العينة الناتج عن التجربة العشوائية، بحيث يحتوي هذا الفراغ على كل الترتيبات الممكنة لمجموعة من العناصر. فقد يكون الاهتمام مثلا هو بعدد طرق ترتيب خمسة رجال حول مائدة مستديرة لعقد اجتماع دوري، أو بعدد طرق الممكنة لاختيار ثلاثة طلبة لتنظيف فصلهم الدراسي أسبوعيا، أو بعدد الطرق التي تمكن كيميائي من تشكيل فرق عمل حول إحدى التجارب المعقدة من بين عدة باحثين. هذه الترتيبات المختلفة يمكن تنفيذها باستخدام مجموعة من الطرق الهامة لعد العناصر تندرج تحت مفهومي **التباديل** و**التوافيق**.

تعريف (16.4): **التباديل** (Permutations):

أ) عدد الطرق الكلية لتباديل (لترتيب) n من العناصر المختلفة هو

$$n! = n(n-1)(n-2)\ldots 2 \times 1$$

وهو ما يعرف **بالمضروب** (Factorial).

ب) عدد الطرق الكلية لتباديل r من العناصر مسحوبة من مجموعة تحوي n عنصر يتم حسابه بالصيغة التالية:

$$P_r^n = \frac{n!}{(n-r)!} \quad , \quad r \leq n$$

حيث الرمز P هو الحرف الأول من كلمة تباديل باللغة الإنجليزية، ويقرأ P_r^n؛ n تباديل r.

---

[1] حيث أن الفئة الخالية Φ هي فئة جزئية من أي فئة.

[2] البعض قد يستخدم أيضا مصطلح **الإحلال** كمرادف للإرجاع.

ونلاحظ أنه إذا كان $r = n$ في التعريف السابق في الفقرة (ب) فإن:

$$P_r^n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n!$$ [1]

وهذا يعني أن الفقرة (أ) في تعريف (16.4) هي حالة خاصة من القانون العام للتباديل في الفقرة (ب) عندما يكون عدد العناصر المطلوب ترتيبها (أو سحبها) مساوي لعدد العناصر الكلية n.

ملاحظة:

1. يجب التنويه إلى أن الفقرة (ب) في التعريف (16.4) تشمل مفهوم سحب أو اختيار r عنصر من n عنصر **بدون تكرار** أو **بدون إرجاع** (Without Replacement) و**باعتبار الترتيب**، والذي يعني أن كل عنصر من الـ n عنصر الكلية لا يمكن اختياره أكثر من مرة واحدة، أو بمعنى أنه لا يمكن اختياره ثم إرجاعه لفراغ العينة ثم الاختيار من جديد.

2. يمكن تحليل قانون المضروب بالصورة؛

$$\begin{aligned} n! &= n(n-1)! \\ &= n(n-1)(n-2)! \\ &= \ldots \\ &= n(n-1)(n-2)\ldots(n-r)! \quad , r \le n \end{aligned}$$

وكذلك قانون التباديل:

$$\begin{aligned} P_r^n = \frac{n!}{(n-r)!} &= \frac{n(n-1)(n-2)\ldots(n-r+1)(n-r)!}{(n-r)!} \\ &= n(n-1)(n-2)\ldots(n-r+1) \quad , \quad r \le n \end{aligned}$$

مثال (6.4): ما هو عدد الطرق الكلية التي يمكن بها تطبيق (ترتيب) تعلم القراءات السبع للقرآن الكريم لسبعة من طلاب العلم الجدد؟.

الحل:

عدد الطرق الكلية لترتيب هذه القراءات هو $7! = 5040$ طريقة. ولاحظ أن أول طريقة من هذه الطرق الكثيرة هي أن يتعلم السبعة طلاب كلهم أول قراءة من القراءات السبع، وثاني طريقة هي أن يتعلم أول ستة من الطلاب أول طريقة من القراءات ويتعلم الطالب السابع ثاني طريقة من القراءات، وهكذا[2] ... .

مثال (7.4): بكم طريقة يمكن اختيار كرتين من صندوق به 5 كرات مرقمة من 1 إلى 5؟ وذلك

أ. بالإرجاع ب. بدون إرجاع.

الحل:

---

[1] لأن مضروب الصفر يساوي الواحد الصحيح، $0! = 1$.

[2] سيتم طرح مثال للمقارنة بين طرق العد، وعرض العناصر المكونة لكل حالة بالتفصيل بعد تناول طريقة التوافيق.

أ. على افتراض أن سحب الكرات تم بالإرجاع وباعتبار الترتيب، فإن عدد الطرق الكلية لاختيار (سحب) كرتين من 5 كرات بالإرجاع هو $n^r = 5^2 = 25$ طريقة.

ب. عدد الطرق الكلية لسحب كرتين من 5 كرات بدون إرجاع وباعتبار الترتيب هو

$$P_r^n = \frac{n!}{(n-r)!} = \frac{5!}{(5-2)!} = 20 \text{ طريقة}$$

ملاحظة: في المثال السابق، (مثال (7.4))، إذا كان الاختيار بالإرجاع وباعتبار الترتيب، وكان المطلوب هو عدد الطرق الكلية لترتيب 5 كرات، (أو عدد طرق اختيار 5 كرات من 5 كرات)، فإن الحل يكون $n! = 5! = 120$ طريقة.

نظرية (4.4): يمكن ترتيب (تبديل) عدد n من العناصر في دائرة باستخدام $(n-1)!$ طريقة كلية.

فمثلا ، يمكن ترتيب أربعة أشخاص على مائدة دائرية للعشاء بـ $6 = (4-1)!$ طرق مختلفة، وذلك لأن المائدة المستديرة ليس لها نقطة بداية يمكن الارتكاز عليها في ترتيب الأشخاص الأول ثم الثاني ... وهكذا، لهذا كان من الضروري "تثبيت" أحد الأشخاص أولا ثم ترتيب الباقي، وعددهم $(n-1)$ شخص باعتبار ترتيب الشخص الأول، وهذا هو السبب في استخدام الصيغة $(n-1)!$ وليس $n!$.

نظرية (5.4): يمكن ترتيب عدد n عنصر مقسمة مسبقا إلى k من المجموعات المتجانسة فيما بينها، بحيث يكون في المجموعة الأولى n_1 عنصر، وفي المجموعة الثانية n_2 عنصر، ... ، وفي المجموعة k يوجد n_k عنصر، (بحيث أن $n_1 + n_2 + \ldots + n_k = n$)، بعدد يساوي

$$\frac{n!}{n_1!\, n_2! \ldots n_k!} \quad \text{طريقة كلية}$$

مثال (8.4): عدد الطرق الكلية لترتيب 5 طلبة من قسم الإحصاء، و 4 طلبة من قسم الرياضيات، و 3 طلبة من قسم الحاسوب في صف معين داخل إحدى القاعات الدراسية لإحدى الكليات العلمية، بحيث يكون طلبة القسم الواحد مع بعضهم البعض هو

$$\frac{n!}{n_1!\, n_2! \ldots n_k!} = \frac{12!}{5!\, 4!\, 3!} = 27720 \text{ طريقة}$$

في بعض التجارب العشوائية، قد يكون المراد هو ترتيب مجموعة من العناصر من بين العناصر الكلية في فراغ العينة وذلك بدون إرجاع وبإهمال الترتيب، في هذه الحالة فإننا نستخدم طريقة عد تعرف بالتوافيق.

تعريف (17.4): **التوافيق** (Combinations): عدد الطرق الكلية لترتيب عدد r عنصر من بين n عنصر بدون إرجاع وبإهمال الترتيب يتم حسابه بالصيغة التالية:

$$C_r^n = \frac{n!}{(n-r)!\, r!} \quad , \quad r \le n$$

ملاحظة: يلاحظ من التعريف السابق أن:

$$P_r^n = C_r^n \; r!$$

وكذلك

$$C_r^{n+1} = C_r^n + C_{r-1}^n$$

مثال (9.4): بكم طريقة يمكن اختيار مجموعة مكونة من 3 رجال و 4 سيدات من بين 5 رجال و 6 سيدات تطوعوا لإجراء دراسة طبية حول الأسباب المساعدة في ارتفاع ضغط الدم.

الحل:

بالنسبة للرجال فإنه يمكن اختيار[1] 3 من 5 رجال بعدد

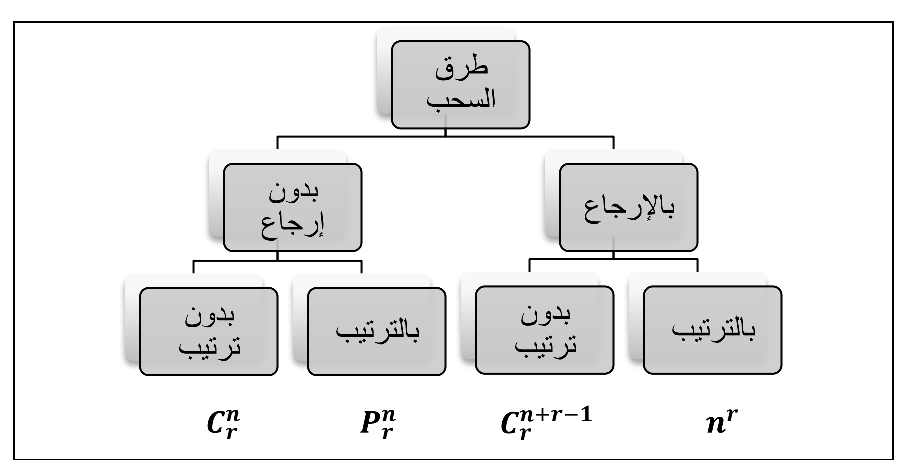
$$C_r^n = \frac{n!}{(n-r)!\,r!} = \frac{5!}{(5-3)!\,3!} = 10 \text{ طرق}$$

وبالنسبة للسيدات فإنه يمكن اختيار 4 من بين 6 سيدة بعدد

$$C_4^6 = \frac{6!}{(6-4)!\,4!} = 15 \text{ طريقة}$$

وبالتالي فإن عدد الطرق الكلية لاختيار المجموعة بالكامل هو 10 × 15 = 150 طريقة.

مما سبق يمكننا تلخيص أساليب العد السابقة، اعتمادا على طريقة السحب أو الترتيب، كما هو موضح في المخطط التالي، (شكل (5.4)):



شكل (5.4): طرق العد المختلفة بناء على أسلوب السحب (الاختيار).

مثال (10.4): مستخدما فراغ العينة $A = \{\, x, y, z \,\}$ أوجد كل العينات الممكن سحبها ذات الحجم 2 من A:

1. بالإرجاع والترتيب. 2. بدون إرجاع وبالترتيب. 3. بدون إرجاع وبإهمال الترتيب.

الحل:

[1] لاحظ أننا عادة ما نستخدم طريقة التوافيق في التجارب أو المسائل التي تتعلق باختيار الأشخاص لأنه من المنطق عدم تكرار الشخص عند الاختيار، وكذلك إهمال ترتيب الأشخاص بينهم البعض.

1. بالإرجاع والترتيب: يمكن في هذه الحالة الحصول على $n^r = 3^2 = 9$ عينات أو طرق لسحب عنصرين أو مفردتين من A. هذه العينات أو الطرق هي:

xx, xy, xz
yx, yy, yz
zx, zy, zz

2. بدون إرجاع وبالترتيب: لدينا

$$P_r^n = \frac{n!}{(n-r)!} = \frac{3!}{(3-2)!} = 6 \text{ طرق}$$

هي

xy, xz, yx
yz, zx, zy

3. بدون إرجاع وبإهمال الترتيب: لدينا

$$C_2^3 = \frac{3!}{(3-2)!\,2!} = 3 \text{ طرق}$$

هي

xy, xz, yz

ومن ضمن استخدامات التوافيق الهامة، استخدامه في مفكوك ذو الحدين، والذي سنتطرق له بتفصيل أوسع في الفصول القادمة عند تناول التوزيعات الاحتمالية.

نظرية (6.4): **مفكوك ذو الحدين** (Binomial Expansion):

لأي عدد صحيح $n \geq 0$ يكون

$$(x+y)^n = \sum_{r=0}^{n} C_r^n \, x^{n-r} y^r \quad , \quad r \leq n$$

فمثلا، عندما يكون $n = 2$ فإن:

$$(x+y)^2 = \sum_{r=0}^{2} C_r^2 \, x^{2-r} y^r$$

$$= C_0^2 x^2 + C_1^2 xy + C_2^2 y^2$$

$$= x^2 + 2xy + y^2$$

في بعض المسائل المتعلقة بطرق العد، قد نتعامل مع مجموعة أو فئة تضم عددا كبيرا من العناصر والتي نحتاج لحساب المضروب لها، وأحيانا نكون بحاجة لتبسيط التعامل مع صيغة رياضية محددة تتضمن استخدام التباديل أو التوافيق، في هذه الحالات، الصيغة التالية يكون لها استخدام مهم:

تعريف (18.4): **صيغة استيرلنج** (Stirling's Formula): إذا كانت $n > 0$ فإن:

$$n! \sim \sqrt{2\pi n}\, n^n e^{-n}$$

حيث

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}\, n^n e^{-n}} = 1$$

وهذه الصيغة تعطي عادة نتائج مقاربة جدا لقيمة مضروب n الحقيقية، كما نرى من جدول (1.4) التالي:

جدول (1.4): نتائج استخدام $n!$ وصيغة استيرلنج مع بعض قيم n.

$\sqrt{2\pi n}\, n^n e^{-n}$	$n!$	n
1.919	2	2
118.019	120	5
475687486.474	479001600	12

ملاحظة:

1. يمكن من خلال التعويض ببعض قيم r في قانوني التباديل والتوافيق الحصول على القيم الخاصة التالية:

عندما			
$r = 0$	$r = 1$	$r = n - 1$	$r = n$
$P_0^n = 1$	$P_1^n = n$	$P_{n-1}^n = n!$	$P_n^n = n!$
$C_0^n = 1$	$C_1^n = n$	$C_{n-1}^n = n$	$C_n^n = 1$

2. يمكن إثبات[1] العلاقات التالية بسهولة، (علما بأن $r \le n$)؛

$$C_r^n < P_r^n < n^r$$

و

$$C_r^n = \frac{P_r^n}{r!}$$

و

$$C_r^{n+1} = C_{r-1}^n + C_r^n$$

و

$$C_r^n = C_{n-r}^n$$

وإذا كان $n_1 + n_2 = n$ فإن

$$C_{n_1}^n = C_{n_2}^n$$

[1] يترك إثبات هذه العلاقات كتمرين للقارئ.

2.3.4 مُسلمات الاحتمال (Probability Axioms)

ذكرنا في البند (1.4) المفهوم المنطقي للاحتمال، ووضحنا من خلال بعض الأمثلة البسيطة كيفية تعامل البعض مع علم الاحتمالات. ثم قمنا بمناقشة الطرق الأساسية لاختيار أو سحب العناصر (العينات) عند إجراء التجارب العشوائية وتكوين فراغ العينة، وهذا بحد ذاته يشكل المكونات الرئيسية التي يعتمد عليها حساب الاحتمال. وسنقوم الآن بعرض تعريف مصطلح "الاحتمال" بصورته الرياضية، ثم نناقش أهم القواعد والخواص المتعلقة به.

تعريف (19.4): **الاحتمال** (Probability): **التعريف التقليدي للاحتمال** (Classical Definition): إذا تم إجراء تجربة عشوائية ما، وكان عدد النواتج الكلية لها هو n حالة متنافية، فإن احتمال حدوث (وقوع) k حالة من n يتم حسابه بالصورة:

$$P(k) = \frac{k}{n} \quad , \quad k \leq n$$

أ. **التعريف الحديث للاحتمال** (Modern Definition): إذا كان S فراغ عينة يحتوي على جميع نتائج التجربة العشوائية، وليكن عددها n، فإنه لأي حدث A معرف على S يكون احتمال الحصول على (وقوع) الحدث A، والذي يرمز له بالرمز $P(A)$، هو عدد العناصر في A مقسوما على عدد العناصر في فراغ العينة، بمعنى:

$$P(A) = \frac{n(A)}{n(S)} \quad , \quad n(A) \leq n(S)$$

تعريف (20.4): **الاحتمال ومسلمات (بديهيات) الاحتمال** (Probability and Probability Axioms): لأي حدث A معرف على فراغ العينة S يكون:

1. $0 \leq P(A) \leq 1$ ، لكل $A \in S$.
2. $P(S) = 1$[1]
3. $P(\Phi) = 0$

مثال (11.4): في تجربة إلقاء عملة معدنية متزنة مرتين يكون لدينا؛ $S = \{HH, HT, TH, TT\}$، فإذا ما عرفنا الأحداث التالية:

A يمثل الحصول على وجه (H) في الرمية الأولى، B يمثل الحصول على كتابة (T) في الرميتين، C يمثل الحصول على صورة وكتابة، و D يمثل عدم الحصول على أي كتابة. فإن احتمال وقوع هذه الأحداث هو

$$P(A) = \frac{n(A)}{n(S)} = \frac{2}{4} = \frac{1}{2} \quad , \quad P(B) = \frac{n(B)}{n(S)} = \frac{1}{4}$$

$$P(C) = \frac{n(C)}{n(S)} = \frac{2}{4} = \frac{1}{2} \quad , \quad P(D) = \frac{n(D)}{n(S)} = \frac{1}{4}$$

[1] حيث أن $P(S) = \frac{n(S)}{n(S)} = 1$.

مثال (12.4): فصل به 30 طالب، منهم 5 طلبة معدلاتهم الدراسية في العام السابق تفوق 85%، تم اختيار أحد الطلبة من هذا الفصل بصورة عشوائية، ما احتمال أن يكون معدله الدراسي أكبر من 85%؟.

الحل:

لنفرض أن A= حدث اختيار طالب معدله الدراسي أكبر من 85%، فيكون: $P(A)=\frac{n(A)}{n(S)}=\frac{5}{30}=\frac{1}{6}$.

مثال (13.4): في إحدى المختبرات العلمية يوجد قفص به 12 فأرا، وكانت 4 من هذه الفئران بيضاء والباقي سوداء. تم اختيار فأرين من القفص بصورة عشوائية لأخذ عينات دم منها، ما احتمال أن تكون هذه الفئران المختارة:

1. جميعها بيضاء؟ 2. أحدها أبيض والآخر أسود؟.

الحل:

1. لنعتبر أن E_1 يمثل حدث اختيار فأرين لونهما أبيض، فيكون لدينا

$$n(E_1)=C_2^4=\frac{4!}{(4-2)!\,2!}= \text{6 طرق للاختيار}$$

وحيث أن عدد الطرق الكلية لاختيار فأرين من 12 فأرا هو

$$n(S)=C_2^{12}=\frac{12!}{(12-2)!\,2!}= \text{66 طريقة}$$

فيكون

$$P(E_1)=\frac{n(E_1)}{n(S)}=\frac{6}{66}=\frac{1}{11}=0.091$$

ولاحظ أن استخدام طريقة التوافيق في اختيار الفئران كان بسبب أن ترتيب الفئران لا أهمية له في هذا المثال، (إلى جانب عدم منطقية الإرجاع في السحب)، بمعنى أن اختيار الفأر الأول والثاني مثلا يكافئ اختيار الفأر الثاني والأول. أما إذا كانت طبيعة التجربة العشوائية (المعملية) تفرض أهمية لترتيب هذه الفئران[1]، فإن الاختيار في هذه الحالة، (مع بقاء الاختيار بدون إرجاع)، يتم باستخدام طريقة التباديل.

2. باعتبار أن E_2 يمثل حدث اختيار فأر أبيض والآخر أسود، يكون

$$n(E_2)=C_1^4C_1^8=4\times 8= \text{32 طريقة للاختيار}$$

وبالتالي يكون

$$P(E_2)=\frac{n(E_2)}{n(S)}=\frac{32}{66}=\frac{16}{33}=0.485$$

---

[1] كأن توضع اعتبارات أخرى للتجربة العشوائية مثل عمر الفأر وحالته الجسدية، أو وجود فروق زمنية محددة عند الاختيار، أو مراقبة تأثر فئران لها تصنيف معين، ...، وغيرها.

3.3.4 **النظريات الأساسية للاحتمال** (Basic Theorems for Probability)

نظرية (7.4): **قاعدة الجمع** (Additive Rule): إذا كان A و B هما أي حدثين معرفين على فراغ العينة S فإن:

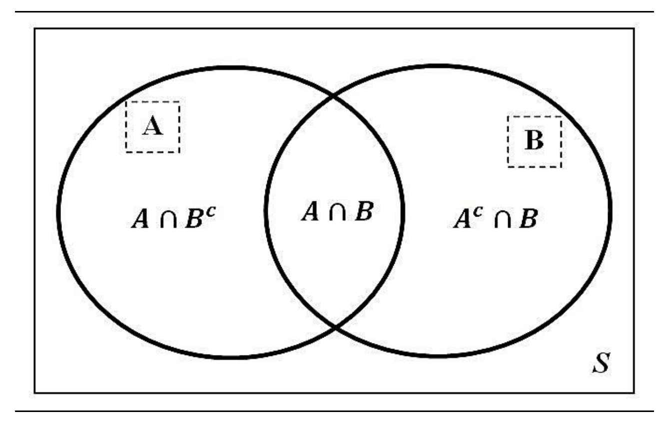
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

الإثبات:

من الشكل (6.4) نلاحظ أن $A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$ ، وحيث أن الأحداث $(A \cap B^c)$، $(A \cap B)$، و $(A^c \cap B)$ هي كلها أحداث متنافية، فإن

$$P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(A^c \cap B)$$

ونلاحظ أيضا من الشكل المرفق أن:



شكل (6.4): تجزئة اتحاد الحدثين A و B.

$$P(A) = P(A \cap B^c) + P(A \cap B)$$

وكذلك

$$P(B) = P(A^c \cap B) + P(A \cap B)$$

إذن

$$\begin{aligned} P(A \cup B) &= P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

نتيجة (1.4): إذا كان A و B حدثين متنافيين في فراغ العينة S فإن:

$$P(A \cup B) = P(A) + P(B)$$

وهذه النتيجة يمكن تعميمها بالصورة التالية:

نتيجة (2.4): لأي مجموعة منتهية من الأحداث المتنافية $A_1, A_2, \ldots, A_k$ يكون:

$$P(\bigcup_{i=1}^{k} A_i) = \sum_{i=1}^{k} P(A_i)$$

فمثلا؛ عند رمي زهر نرد يكون فراغ العينة $S = \{1, 2, 3, 4, 5, 6\}$ وبالتالي إذا ما وضعنا الأحداث؛ ظهور الرقم 1 = A_1، ظهور الرقم 2 = A_2، ... ، ظهور الرقم 5 = A_5. فإن

$$P(\bigcup_{i=1}^{5} A_i) = \sum_{i=1}^{5} P(A_i) = 5 \times \frac{1}{6} = 0.833$$

مثال (14.4): في إحدى مباريات كرة القدم، يلعب الفريق الأول بقمصان زرقاء ويلعب الفريق الثاني بقمصان خضراء. ويلبس 4 لاعبين من الفريق الأول الأحذية الرياضية السوداء، ومن الفريق الثاني يلبس 5 لاعبين الأحذية الرياضية البيضاء[1]. فإذا ما تم:

أ. اختيار لاعب واحد من الملعب بصورة عشوائية، فما احتمال أن يكون:
1. من الفريق الأول؟.
2. مرتديا لحذاء أسود؟.
3. من الفريق الأول ومرتديا لحذاء أسود؟.
4. من الفريق الأول أو مرتديا لحذاء أسود؟.

ب. اختيار لاعبين اثنين من الملعب عشوائيا، ما احتمال أن يكون كلا اللاعبين:
1. من الفريق الأول؟.
2. من الفريق الأول أو يرتدون أحذية سوداء؟.

الحل:

من المعلوم أن الفريق الواحد في مباراة كرة القدم يضم 11 لاعبا على أرض الملعب، وبالتالي يكون:

أ.

1. لنضع A = حدث أن اللاعب هو من الفريق الأول، عندئذ يكون

$$P(A) = \frac{n(A)}{n(S)} = \frac{11}{22} = 0.167$$

2. بوضع B = حدث أن اللاعب يرتدي الحذاء الأسود، فيكون

$$P(B) = \frac{n(B)}{n(S)} = \frac{10}{22} = 0.455$$

3. في هذه الحالة، يتمتع اللاعب المختار بصفتين؛ أن يكون من الفريق الأول ويرتدي حذاء أسود. وهكذا يكون

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{4}{22} = 0.182$$

---

[1] على افتراض أن كل اللاعبين في الفريقين إما أن يرتدوا الأحذية السوداء أو البيضاء.

4. باستخدام قاعدة الجمع نحصل على

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$P(A \cup B) = \frac{1}{2} + \frac{5}{11} - \frac{2}{11} = \frac{17}{22} = 0.773$$

ب.

1. لنضع C حدث أن كلا من اللاعبين من الفريق الأول، فيكون

$$P(C) = \frac{n(C)}{n(S)} = \frac{C_2^{11}}{C_2^{22}} = \frac{55}{231} = 0.238$$

2. بوضع D = حدث أن كلا اللاعبين يرتديان الحذاء الأسود، نحصل على

$$P(D) = \frac{n(D)}{n(S)} = \frac{C_2^{10}}{C_2^{22}} = \frac{45}{231} = 0.195$$

ولحساب الاحتمال المطلوب، وهو $P(C \cup D)$، نحتاج أولا لحساب الاحتمال $P(C \cap D)$؛

$$P(C \cap D) = \frac{n(C \cap D)}{n(S)} = \frac{C_2^{4}}{C_2^{22}} = \frac{6}{231} = 0.026$$

وبالتالي احتمال أن يكون كلا اللاعبين إما من الفريق الأول أو يرتديان الأحذية السوداء هو

$$P(C \cup D) = P(C) + P(D) - P(C \cap D)$$
$$= \frac{55}{231} + \frac{45}{231} - \frac{6}{231} = \frac{94}{231} = 0.407$$

وقاعدة الجمع المعرفة في النظرية (7.4) يمكن تعميمها بالصورة التالية:

نظرية (8.4): **قاعدة الجمع العامة** (Generalized Additive Rule): إذا كانت الأحداث $A_1, A_2, ..., A_k$ هي أحداث معرفة على فراغ العينة S، فإن:

$$P(\bigcup_{i=1}^{k} A_i) = P(A_1 \cup A_2 \cup ... \cup A_k)$$
$$= \sum_{i=1}^{k} P(A_i)$$
$$- \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} P(A_i \cap A_j)$$
$$+ \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \sum_{l=j+1}^{k} P(A_i \cap A_j \cap A_l) - \cdots + (-1)^{k-1} P(A_1 \cap A_2 \cap ... \cap A_k)$$

ملاحظة: كحالة خاصة من النظرية السابقة، عندما نتعامل مع $k = 3$ أحداث يكون

$$P(A_1 \cup A_2 \cup A_3)$$
$$= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3)$$
$$- P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$$

نظرية (9.4): لأي حدث A معرف على فراغ العينة S يكون $P(A^c) = 1 - P(A)$

الإثبات:

حيث أن A و A^c هي أحداث متنافية، وحيث أن $A \cup A^c = S$، فإنه يكون

$$P(S) = P(A) + P(A^c) = 1$$

وبالتالي

$$P(A^c) = 1 - P(A)$$

فمثلا؛ في تجربة إلقاء زهر نرد وعملة معدنية يكون

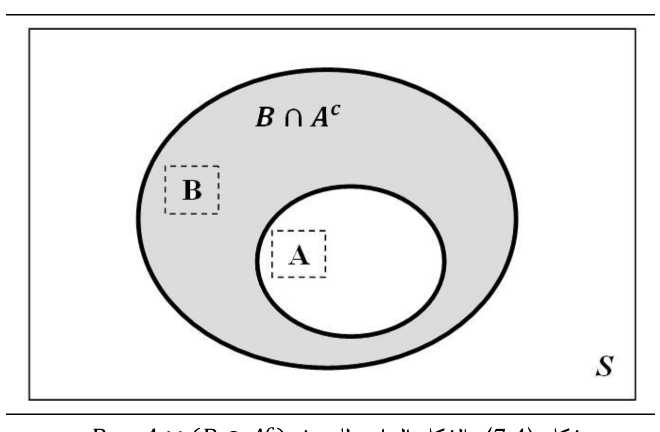
S = {H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6}، فإذا ما عرفنا الحدث A بأنه يمثل ظهور الصورة، فإنه يكون $P(A) = \frac{6}{12} = \frac{1}{2}$.

ويكون الحدث A^c هو الحدث الذي يضم كل الحالات التي لا تظهر فيها الصورة فيكون $P(A^c) = \frac{6}{12} = \frac{1}{2}$، ومن الواضح أن $P(A^c) = 1 - P(A) = 1 - \frac{1}{2} = \frac{1}{2}$.

نظرية (10.4): إذا كان A و B هما حدثين معرفين على فراغ العينة S، وكان $A \subset B$ ، (أي أن A هو فئة جزئية من B)، فإن $P(A) \leq P(B)$.

الإثبات:

كما يوضح لنا الشكل (7.4)، فإن الحدث B يمكن كتابته كالتالي $B = A \cup (B \cap A^c)$ وهكذا يكون $P(B) = P(A) + P(A \cap A^c)$ وحيث أن $P(B \cap A^c) \geq 0$ فإن العلاقة (المتباينة) $P(A) \leq P(B)$ تكون صحيحة.



شكل (7.4): الشكل البياني للحدث $B = A \cup (B \cap A^c)$.

فمثلا؛ في تجربة إلقاء عملة معدنية متزنة مرتين يكون S = {HH, HT, TH, TT}، فإذا ما عرّفنا الحدث B بأنه يمثل ظهور صورة واحدة على الأقل في الرميتين، فإن B = {HH, HT, TH} ، وإذا ما عرّفنا الحدث A بأنه يمثل ظهور صورتين بالضبط، فإن A = {HH}، وعليه فإن $A \subset B$ وبالتالي يكون $P(A) = \frac{1}{4} < P(B) = \frac{3}{4}$.

نظرية (11.4): إذا كان A و B حدثين معرّفين على فراغ العينة S فإن:

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

الإثبات:

حيث أن

$$A = A \cap S = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c)$$

وبما أن كلا من $(A \cap B)$ و $(A \cap B^c)$ هي أحداث متنافية فإن

$$P(A) = P\big((A \cap B) \cup (A \cap B^c)\big) = P(A \cap B) + P(A \cap B^c)$$

مثال (15.4): في أحد الأقسام العلمية لكلية الهندسة، أراد رئيس القسم تكوين لجنة علمية لتقييم أطروحة ماجستير مقدمة من أحد طلبة الدراسات العليا في القسم. وكانت لجان المناقشة تشكل عادة من أربعة من أعضاء هيئة تدريس، وحيث أنه يوجد بالقسم 9 أعضاء من الرجال و 3 من السيدات كلهم مؤهلون للمشاركة في التقييم، فما احتمال أن يكون ضمن هذه اللجنة العلمية:

1. رجلان؟ 2. ثلاثة سيدات؟ 3. سيدة واحدة على الأقل؟.

الحل:

لدينا $n(S) = C_4^{12} = 495$ طريقة كلية لاختيار 4 أعضاء من كل الأشخاص المؤهلين. ولنرمز لأعضاء هيئة التدريس الرجال بالرمز m والسيدات بالرمز f فيكون

.1

$$P(m,m) = \frac{C_2^9 \times C_2^3}{C_4^{12}} = \frac{108}{495} = 0.218$$

.2

$$P(f,f,f) = \frac{C_3^3 \times C_1^9}{C_4^{12}} = \frac{9}{495} = 0.018$$

.3

$$P(f \geq 1) = P(f) + P(f,f) + P(f,f,f) = \frac{C_1^3 \times C_3^9 + C_2^3 \times C_2^9 + C_3^3 \times C_1^9}{C_4^{12}} = \frac{369}{495}$$
$$= 0.746$$

4.4 الاحتمال الشرطي ونظرية بيز (Conditional Probability and Bayes Theorem)

في كثير من الأحيان، وخلال تعاملنا مع الاحتمالات المترابطة في التجربة أو الظاهرة الواحدة، قد نكون بصدد حساب احتمال وقوع حدث معين، (وليكن الحدث A)، عند توفر "معلومات" عن وقوع حدث آخر، (وليكن B)، قبل حدوث A، علما بأن كلا الحدثين هما لنفس التجربة العشوائية ومعرّفان على نفس فراغ العينة.

في هذه الحالة، فإن احتمال وقوع هذه التركيبة من الأحداث يسمى بالاحتمال الشرطي لوقوع الحدث A علما بأن (أو بشرط أن) الحدث B قد وقع مسبقا، ويرمز له بالرمز $P(A\backslash B)$.

تعريف (21.4): **الاحتمال الشرطي** (Conditional Probability): إذا كان كلا من A و B حدثان معرّفان على فراغ العينة S فإن الاحتمال الشرطي لوقوع الحدث A علما بأن الحدث B قد وقع مسبقا يعرّف بالصورة

$$P(A\backslash B) = \frac{P(A \cap B)}{P(B)} \quad , \quad P(B) > 0$$

مثال (16.4): الجدول المرفق يوضح توزيع أعداد الطلبة في أحد أقسام كلية العلوم حسب الجنس والمستوى الدراسي. فإذا ما تم تعريف الأحداث التالية على فراغ العينة:

A_1 يمثل حدث اختيار طالبة من القسم، A_2 يمثل حدث اختيار شخص (طالب أو طالبة) مستواه الدراسي جيد، A_3 يمثل حدث اختيار شخص مستواه الدراسي مقبول، و A_4 يمثل حدث اختيار طالب من القسم.

		المستوى الدراسي		
		مقبول	جيد	جيد جدا فأكثر
الجنس	ذكر	50	40	25
	أنثى	90	60	35

فإنه يمكن حساب الاحتمالات التالية:

1. احتمال اختيار طالبة من القسم علما بأن (بشرط أن) مستواها الدراسي جيد هو
$$P(A_1\backslash A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)} = \frac{60/300}{(40+60)/300} = 0.60$$
حيث أن $n(S) = 300$، والذي يمثل مجموع كل الطلاب بكافة المستويات الدراسية.

2. احتمال اختيار شخص مستواه الدراسي مقبول علما بأنه طالب من القسم هو
$$P(A_3\backslash A_4) = \frac{P(A_3 \cap A_4)}{P(A_4)} = \frac{50/300}{(50+40+25)/300} = 0.43$$

3. احتمال اختيار طالب أو طالبة من القسم هو
$$P(A_4 \cup A_1) = P(A_4) + P(A_1) = \frac{115}{300} + \frac{185}{300} = 1$$
حيث أن $A_4 = A_1^c$

4. احتمال اختيار طالبة أو شخص مستواه الدراسي جيد هو
$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = \frac{185}{300} + \frac{100}{300} - \frac{60}{300} = 0.75$$
ولاحظ أن هذا الاحتمالات المحسوبة في (3) و (4) ليست احتمالات شرطية.

نظرية (12.4): إذا كان S هو فراغ العينة لتجربة عشوائية ما، وكان B هو حدث معرّف على S حيث $P(B) > 0$، فإن:

1. $P(A\backslash B)\geq 0$ لأي حدث A في S.
2. $P(S\backslash B)=1$.
3. إذا كانت $A_1, A_2, A_3, \ldots$ هي متتابعة من الأحداث المتنافية في S فإن

$$P(\bigcup_{i=1}^{\infty} A_i\backslash B)=\sum_{i=1}^{\infty} P(A_i\backslash B)$$

تعريف (22.4): **الأحداث المستقلة والاحتمال الشرطي**

(Independent Events and Conditional Probability):

إذا كان A و B حدثان معرّفان على فراغ العينة S، فإنهما يكونا مستقلين إذا كان

$$P(A\backslash B)=P(A) \text{ أو } P(B\backslash A)=P(B)$$

حيث أنه في حالة الاستقلال بين الأحداث A و B يكون $P(A\cap B)=P(A)P(B)$.

مثال(17.4): في تجربة إلقاء عملة معدنية مرتين يكون S[1] $=\{H_1H_2, H_1T_2, T_1H_2, T_1T_2\}$، وعندها يمكننا مثلا حساب احتمال الحصول على صورة في الرمية الأولى بشرط الحصول على كتابة في الرمية الثانية:

$$P(H_1\backslash T_2)=\frac{P(H_1\cap T_2)}{P(T_2)}=\frac{P(H_1)P(T_2)}{P(T_2)}=P(H_1)=\frac{1}{2}$$

بمعنى أن الحصول على صورة في الرمية الأولى لا يعتمد (مستقل) عن الحصول على كتابة في الرمية الثانية، (ومستقل أيضا عن حدث الحصول على صورة في الرمية الثانية).

تعريف (23.4): **قاعدة الضرب للاحتمال الشرطي**

(Multiplication Rule for Conditional Probability):

إذا كان A و B هما أي حدثين في فراغ العينة S فإن

$$P(A\cap B)=P(A)P(B\backslash A)=P(B)P(A\backslash B)$$

ويمكن تعميم التعريف السابق بالصورة:

إذا كان $A_1, A_2, \ldots, A_k$ هي أحداث معرّفة على فراغ العينة S فإن:

$$P(A_1\cap A_2\cap \ldots\cap A_k)$$
$$=P(A_1)P(A_2\backslash A_1)P(A_3\backslash A_1\cap A_2)\ldots P(A_k\backslash A_1\cap A_2\cap \ldots\cap A_{k-1})$$

وهي ما تعرف **بقاعدة الضرب العامة للاحتمال الشرطي** (Generalized Multiplication Rule for Conditional Probability).

[1] ترقيم نتائج الصورة والكتابة هو فقط لتركيز الانتباه إلى ترتيب الرمية.

مثال (18.4): صندوق به 8 كرات حمراء و 4 كرات بيضاء. تم سحب 3 كرات عشوائيا من هذا الصندوق الواحدة تلو الأخرى بدون إرجاع. أوجد احتمال أن تكون:

1. الكرتان الأولى والثانية حمراء والثالثة بيضاء. 2. جميع الكرات حمراء.

الحل:

إذا ما رمزنا للكرات الحمراء بالرموز $R_1, R_2, \dots, R_8$، والكرات البيضاء بالرموز W_1, W_2, W_3, W_4 فإنه يمكن حساب المطلوب بالصورة:

1. $P(R_1 \cap R_2 \cap W_3) = P(R_1)P(R_2 \backslash R_1)P(W_3 \backslash R_1 \cap R_2)$

حيث

$$P(R_1) = \frac{8}{12}$$

$$P(R_2 \backslash R_1) = \frac{P(R_1 \cap R_2)}{P(R_1)} = \frac{\frac{8}{12} \times \frac{7}{11}}{\frac{8}{12}} = \frac{7}{11}$$

$$P(W_3 \backslash R_1 \cap R_2) = \frac{4}{10}$$

وبالتالي يكون:

$$P(R_1 \cap R_2 \cap W_3) = \frac{8}{12} \times \frac{7}{11} \times \frac{4}{10} = \frac{28}{165} = 0.776$$

2.

$$P(R_1 \cap R_2 \cap R_3) = P(R_1)P(R_2 \backslash R_1)P(R_3 \backslash R_1 \cap R_2) = \frac{8}{12} \times \frac{7}{11} \times \frac{6}{10} = \frac{14}{55}$$
$$= 0.255$$

نظرية (13.4): **قانون مجموع الاحتمالات** (Law of Total Probability):

إذا كان B هو أي حدث في فراغ العينة S، حيث $P(B) > 0$ و $P(B^c) > 0$، فإنه لأي حدث A يكون:

$$P(A) = P(A \backslash B)P(B) + P(A \backslash B^c)P(B^c)$$

الإثبات:

لدينا من النظرية (11.4) العلاقة $P(A) = P(A \cap B) + P(A \cap B^c)$، وحيث أنه من التعريف (23.4) لدينا $P(A \cap B) = P(B)P(A \backslash B)$، وكذلك يكون $P(A \cap B^c) = P(B^c)P(A \backslash B^c)$، عندئذ بالتعويض في الطرف الأيمن المقابل للاحتمال $P(A)$ نحصل على

$$.P(A) = P(A \backslash B)P(B) + P(A \backslash B^c)P(B^c)$$

و تعميم النظرية السابقة بحيث يتم التعامل مع كل الأحداث $B_1, B_2, ..., B_k$ والتي تشكل تقسيم لفراغ العينة S يشكل في الواقع الأساس للنظرية التي سنستعرضها فيما يلي، والتي تعد من نظريات الاحتمال التي يرتكز عليها أحد أهم فروع علم الاحتمال بل وعلم الإحصاء أيضا[1].

نظرية (14.4): **نظرية بيز** (Bayes Theorem):

إذا كانت $B_1, B_2, ..., B_k$ هي أحداث تمثل تقسيم لفراغ العينة S، حيث $P(B_i) \neq 0$ لكل $i = 1, 2, \dots, k$ ، فإنه لأي حدث A في S ، بحيث $P(A) \neq 0$، يكون احتمال وقوع الحدث B_j ، $(B_j \in S)$، بشرط وقوع الحدث A يعرّف بالصورة:

$$P(B_j \backslash A) = \frac{P(B_j)P(A \backslash B_j)}{P(A)}$$

حيث أن

$$P(A) = P(B_1)P(A \backslash B_1) + P(B_2)P(A \backslash B_2) + \cdots + P(B_k)P(A \backslash B_k)$$

الإثبات:

من تعريف الاحتمال الشرطي لدينا

$$P(B_j \backslash A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(B_j)P(A \backslash B_j)}{P(A)}$$

وذلك لكل $B_j \in S$.

مثال (19.4): في إحدى مصانع السيارات الألمانية، كان هنالك ثلاثة خطوط إنتاج خاصة بتركيب المحركات ذات الستة اسطوانات وهي B_1، B_2، و B_3 . وتقوم هذه الخطوط بتجهيز 50%، 30%، و 20% من عدد السيارات الكلي بالمحركات في المصنع على التوالي[2]. تم اكتشاف عيوب طفيفة في طريقة تركيب هذه المحركات بعد فترة من الإنتاج بحيث كانت نسب العيوب في خطوط التركيب هي 3%، 4%، و 5% على التوالي. فإذا ما تم اختيار سيارة من إنتاج هذا المصنع بطريقة عشوائية بغرض اختبار محركها، ما هو احتمال:

1. أن توجد عيوب في تركيب محركها؟.
2. أن تكون هذه السيارة قد تم تركيب محركها في الخط B_1 إذا علمت بوجود عيوب في تركيب هذا المحرك؟.

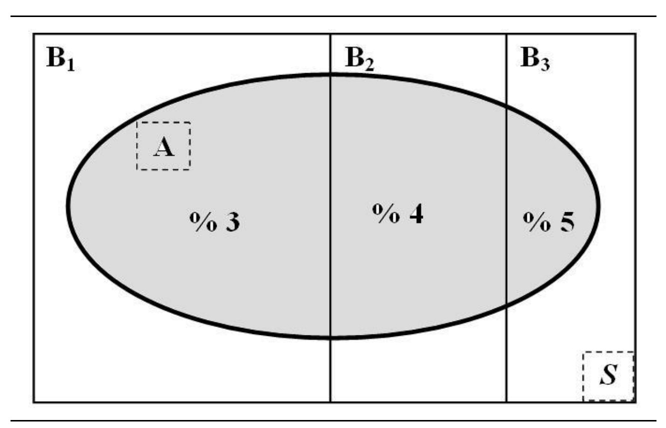
الحل:

1. لنفرض أن الحدث A يمثل أن تكون السيارة المختارة محركها به عيوب. عندئذ يكون لدينا، (كما هو موضح في الشكل (8.4))، الاحتمال التالي:

[1] يوجد توجه في استخدام مفهوم نظرية بيز في الاستدلال الإحصائي يعرف بمدرسة بيز (Bayesian School).
[2] تجدر الإشارة هنا إلى أن الاحتمال إذا ما تم ضرب قيمته بـ 100 فإنه يصبح نسبة مئوية.

$$P(A) = P(B_1)P(A\backslash B_1) + P(B_2)P(A\backslash B_2) + P(B_3)P(A\backslash B_3)$$
$$= 0.50 \times 0.03 + 0.30 \times 0.04 + 0.20 \times 0.05 = 0.037$$

وهذا يعني أن احتمال وجود أو نسبة العيوب في محرك أي سيارة منتجة في هذا المصنع هي 3.7%.



شكل (8.4): تقسيم خطوط الإنتاج في المثال (19.4).

2. بتطبيق نظرية بيز نحصل على

$$P(B_1\backslash A) = \frac{P(B_1)P(A\backslash B_1)}{P(A)} = \frac{0.50 \times 0.03}{0.037} = 0.41$$

مما يعني أن احتمال أن يكون محرك السيارة المسحوبة تم تركيبه في الخط B_1 علما بأنه معيب هو 0.41 ، أي بنسبة 41%.

4.5 تمارين الفصل الرابع

تمرين (1.4): إذا قام أحد الطلبة بدراسة مقرر دراسي واحد في الرياضيات في كل فصل دراسي. وعلى فرض أنه أنهى ثلاث فصول دراسية، فأوجد كل النتائج الممكن أن يكون قد تحصل عليها هذا الطالب في هذه المقررات الرياضية، علما بأن النتائج الممكنة لأي مقرر هي D, C, B, A و F.

تمرين (2.4): فراغ العينة التالي يحوي أسماء مقررات تُدرس في كلية العلوم ويشترك فيها الطلبة من كل الأقسام في الكلية؛ { رياضة1، رياضة2، كيمياء1، نبات1، نبات2، إحصاء1، إحصاء2، فيزياء1، فيزياء2، لغة إنجليزية } = S . وتم تعريف الأحداث التالية والتي تمثل المقررات التي يدرسها الطلاب في بعض الأقسام:

قسم الرياضيات = {رياضة1، رياضة2، إحصاء1، لغة إنجليزية}، قسم الإحصاء = {رياضة1، إحصاء1، إحصاء2، لغة إنجليزية}، قسم الكيمياء = {كيمياء1، لغة إنجليزية}. أوجد كل العلاقات التي وردت في الجزء (3.2.4) في الكتاب بين هذه الأحداث.

تمرين (3.4): كم تركيبة يمكن تكوينها من المعادن الكيميائية التالية: { زنك، نحاس، فضة، ألومنيوم، كوبالت، حديد }؟.

تمرين (4.4): بكم طريقة يمكن اختيار 3 كتب من صندوق به 10 كتب وذلك:

أ. بالإرجاع (واعتبار الترتيب)، ب. بدون إرجاع مع اعتبار الترتيب، ج. بدون إرجاع وبإهمال الترتيب.

تمرين (5.4): بكم طريقة يمكن شراء سيارتين من نوع هيونداي من معرض سيارات به 3 سيارات هيونداي بيضاء و 4 سيارات هيونداي سوداء؟.

تمرين (6.4): بكم طريقة يمكن ترتيب 4 أطفال، 3 بالغين، و 4 مسنين لأخذ صورة تذكارية بحيث يكون الأشخاص من نفس الفئة العمرية إلى جانب بعضهم البعض؟.

تمرين (7.4): بكم طريقة يمكن لمجموعة من الأشخاص اختيار نوعين من 4 أنواع من الحساء، و 3 أنواع من 6 أطباق من المقبلات، و 3 أنواع من 5 أطباق من اللحوم ضمن مائدة عشاء مفتوح (بوفيه)؟.

تمرين (8.4): يوجد في قسم أمراض الدم في أحد المستشفيات 25 مريض، 5 منهم لديهم فقر دم حاد. تم اختيار أحد المرضى من هذا القسم عشوائيا، ما احتمال:

أ. أن يكون من اللذين لديهم فقر دم حاد؟ ب. أن لا يكون من اللذين لديهم فقر دم حاد؟.

تمرين (9.4): أحد محلات بيع الهواتف النقالة لديه فقط 5 هواتف آيفون 6 بيضاء، و 6 هواتف جالاكسي نوت 5 بيضاء. أرادت إحدى السيدات شراء هاتفين من المحل، ما احتمال أن يكون:

أ. أحدهما آيفون 6 والآخر نوت 5؟ ب. كلاهما نوت 5؟ ج. كلاهما من نفس النوع؟.

تمرين (10.4): أحد محلات بيع مكيفات الهواء به نوعين من المكيفات؛ نظام القطعة الواحدة وعددها 10، ونظام القطعتين وعددها أيضا 10. ويوجد 6 مكيفات من نظام القطعة الواحدة قوتها 24 BTU، و 7 مكيفات من نظام القطعتين قوتها 12 BTU. وتم اختيار مكيفين عشوائيا من المحل، ما احتمال أن يكون كلا المكيفين:

أ. من نظام القطعة الواحدة؟ ب. قوته 24؟ ج. من نظام القطعة الواحدة و قوته 24؟ د. من نظام القطعة الواحدة أو قوته 24؟.

تمرين (11.4): أراد أحد مهندسي الديكور تأثيث منزل مكون من 5 غرف نوم، وكان عليه اختيار الأثاث الخاص بالمنزل من معرضين للأثاث؛ المعرض 1 وبه 10 غرف نوم مختلفة، والمعرض 2 وبه 5 غرف نوم مختلفة. ما احتمال أن يتم تأثيث:

أ. غرفتين من المعرض 1، ب. 3 غرف من المعرض 2، ج. غرفة واحدة على الأقل من المعرض 2، د.غرفتين على الأكثر من المعرض 2.

تمرين (12.4): باستخدام بيانات تمرين (10.4) أعلاه أوجد احتمال اختيار:

أ. مكيف قوته 12 بشرط أن يكون من نظام القطعتين.

ب. مكيف من نظام القطعة الواحدة بشرط أن تكون قوته 24.

ج. مكيف قوته 12 أو من نظام القطعتين.

تمرين (13.4): أنجبت إحدى الأسر 3 مرات.

أ. اكتب فراغ العينة لهذه التجربة.

ب. ما احتمال الحصول على مولود ذكر في الإنجاب الثاني بشرط الحصول على مولودة انثى في الإنجاب الأول.

تمرين (14.4): سلة فاكهة بها 9 ثمرات من التفاح و 5 ثمرات من البرتقال. تم اختيار 3 ثمرات عشوائيا من السلة الواحدة تلو الأخرى بدون إرجاع، أوجد احتمال أن تكون:

أ. جميع الثمرات المسحوبة من التفاح ب. الثمرة الأولى من التفاح والباقي من البرتقال.

تمرين (15.4): ثلاثة صناديق تحوي زجاجات مياه غازية عبوة 330 ملليتر، 500 ملليتر، و1000 ملليتر. وكانت نسبة الزجاجات منتهية الصلاحية هي 25% ، 20% ، و 15% على التوالي. تم اختيار زجاجة عشوائيا من أحد الصناديق، أوجد احتمال أن تكون:

أ. منتهية الصلاحية ب. قد تم اختيارها من صندوق ال 330 ملليتر علما بأنها منتهية الصلاحية.

الفصل الخامس

المتغيرات العشوائية والتوزيعات الاحتمالية
(Random Variables and Probability Distributions)

1.5 **مقدمة** (Introduction)

تناولنا في ما سبق كيفية التعامل مع التجارب العشوائية المختلفة، وطريقة قراءة نتائج هذه التجارب وتنظيمها في فراغ العينة. ولاحظنا أنه في معظم تلك التجارب العشوائية كانت النتائج ممثلة بأرقام بسيطة تكفي لتوضيح مفهوم تلك التجارب، إلا أنه في الكثير من التجارب العملية قد يكون عدد العناصر في فراغ العينة كبير جدا بحيث يصعب التعامل معها مباشرة، أو قد نكون مهتمين "بسلوك" حدث أو مجموعة محددة من الأحداث ضمن فراغ العينة ولا يلزمنا رصد أو تسجيل كل نتائج فراغ العينة. أضف إلى ذلك أن تعاملنا مع نتائج التجارب العشوائية إلى الآن كان مقتصرا على النتائج المنفصلة، أي تلك النتائج التي يمكن رصدها بأعداد منفصلة في فراغ العينة، ولكن كيف يكون الحال إذا ما كانت قيم نتائج التجارب متصلة ولا يمكن التعبير عنها بأعداد منفصلة؟، بمعنى أن يكون ناتج التجربة فترة متصلة من النقاط.

للتعامل مع الحالات السابقة، نجد أنفسنا بحاجة لاستخدام طريقة أكثر مرونة وسهولة في التعامل مع نتائج التجارب العشوائية من أجل تنظيم هذا التعامل بصورة رياضية، ولهذا نجد أنه من الضروري استخدام مفهوم المتغيرات لتنظيم نتائج التجارب. وحيث أنه لا يمكن تحديد أي من هذه النتائج سوف يتم الحصول عليه "بالضبط" ضمن كل نتائج التجربة، فإن هذه المتغيرات تسمى **متغيرات عشوائية**، والتي سيتم تعريفها في البند التالي، ثم سنقوم بتوضيح أنواعها (منفصلة ومتصلة)، وكذلك طرق ارتباطها بالاحتمالات (كتوزيعات احتمالية)، وطرق التعامل مع دوالها بشكل رياضي.

2.5 **المتغير العشوائي** (Random Variable)

لنفرض أنه لدينا المتغير[1] X والذي يأخذ القيم $X = \{1, 2, 3, 4, 5\}$ ، وتم تعريف الدالة $f(x)$ بأنها تساوي مربع X ، أي أن $f(x) = x^2$ ، (حيث $0 \le x \le 5$)، وهذا يعني أن الدالة $f(x)$ تقوم بتعيين قيمة جديدة لكل قيمة من قيم المتغير X، فعندما يكون $x = 2$ مثلا فإن $f(2) = 4$ وهكذا بالنسبة لكل قيم[2] X، فنلاحظ أن الدالة الرياضية $f(x)$ يمكن أن تأخذ "أي" قيمة عند التعويض بقيم x ولا يوجد قيود محددة عليها كما سنرى عند تعريف الدالة الاحتمالية.

وبغض النظر عن طبيعة النتائج التي يتم الحصول عليها من التجارب العشوائية من حيث كونها رقمية (مثل تجربة رمي زهر نرد)، أو غير رقمية (عند رمي عملة معدنية)، فإن المتغير العشوائي في هذه الحالة سيقوم بتعيين أعداد لكل عنصر من عناصر فراغ العينة الناتج عن التجربة العشوائية. هذا التعيين يمثل في الواقع قاعدة أو قانون يتم وضعه لتنظيم نتائج التجربة، لهذا يمكننا القول أن المتغير العشوائي هو بحد ذاته دالة، وسيتم تعريفه بالصورة التالية:

تعريف (1.5): **المتغير العشوائي** (Random Variable): المتغير العشوائي هو دالة تقوم بتعيين قيم حقيقية لكل عنصر من عناصر فراغ العينة الناتج عن التجربة العشوائية.

---

[1] لاحظ أننا لم نذكر هنا أن X هو متغير عشوائي.

[2] يتم عادة استخدام الأحرف الإنجليزية الكبيرة للدلالة على اسم المتغير، والأحرف الصغيرة للدلالة على قيم ذلك المتغير.

وهذا يعني أن المتغير العشوائي هو دالة نطاقها فراغ العينة، ولاحظ أنه يمكن تعريف عدد كبير من المتغيرات العشوائية على التجربة العشوائية الواحدة أو فراغ العينة الواحد، وذلك بحسب ما تمثله دالة المتغير.

مثال (1.5): في تجربة إلقاء عملة معدنية ثلاث مرات يكون فراغ العينة الناتج هو

$$S = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$$

وإذا ما تم تعريف المتغير العشوائي X بأنه يمثل عدد الصور (H) الظاهرة في كل حالة فإن X سيأخذ القيمة $x = 3$ في حالة الحصول على HHH ويأخذ القيمة $x = 2$ في حالة الحصول على HHT أو HTHأو THH، وهكذا. لذلك يتم تلخيص كل القيم التي يأخذها المتغير X في جدول عادة، كما هو موضح في جدول (1.5). ويمكن تعريف متغير عشوائي آخر، Y مثلا، بحيث يمثل الفرق بين عدد الصور الظاهرة (H) وعدد الكتابات (T) في كل حالة، بمعنى أن:

جدول (1.5): قيم المتغيرات العشوائية المعرفة في مثال (1.5).

فراغ العينة	X	Y
HHH	3	3
HHT	2	1
HTH	2	1
THH	2	1
TTH	1	−1
THT	1	−1
HTT	1	−1
TTT	0	−3

(عدد الكتابات − عدد الصور = Y)، فتكون قيم Y كما هو موضح في جدول (1.5). ومثلا، يمكننا ملاحظة أن هنالك ثلاث حالات يأخذ فيها المتغير العشوائي X القيمة 2 وهذا يعني أنه توجد ثلاث حالات ظهرت فيها صورتان بالضبط، وحالة يأخذ فيها المتغير العشوائي Y القيمة (3−) وحالة أخرى يأخذ فيها القيمة 3 ، بمعنى أنه توجد حالتان تختفي فيهما الكتابات (الحالة الأولى)، وتختفي الصور (الحالة الثانية).

ويوجد نوعان من المتغيرات العشوائية (بحسب طبيعة التجربة العشوائية أو فراغ العينة)؛ النوع الأول هو **المتغير العشوائي المنفصل** أو **المتقطع** (Discrete Random Variable)، وهو المتغير الذي يأخذ قيما قابلة للعد بصورة منفصلة سواء كانت منتهية أو غير منتهية. والمثال السابق (1.5) ينتمي لهذا النوع من المتغيرات.

والنوع الثاني من المتغيرات هو **المتغير العشوائي المتصل** أو **المستمر** (Continuous Random Variable) ، وهو المتغير الذي تكون قيمه معرّفة على فترة مستمرة نقاطها غير قابلة للعد بصورة منفصلة سواء كانت هذه الفترة منتهية (مغلقة) أو غير منتهية (مفتوحة)، كأن يمثل المتغير فترات عمرية أو زمنية ... . وهذا النوع من المتغيرات سيتم تناوله في الجزء التالي.

3.5 التوزيعات الاحتمالية (Probability Distributions)

عرّفنا في البند السابق المتغير العشوائي، وبينّا كيف أنه يمثل دالة تصف أو تعين قيم لنتائج التجربة العشوائية، وفي هذا البند سنوضح كيف أنه إذا ما تم تعيين قيمة احتمالية لكل قيمة من قيم المتغير العشوائي فإن الصورة أو التركيبة الناتجة عن ذلك تعرف بالتوزيع الاحتمالي للمتغير العشوائي. وطبيعة المتغير العشوائي، من

حيث كونه منفصلا أو متصلا، ستنعكس بالطبع على طبيعة التوزيع الاحتمالي، وسنبدأ مع النوع الأول من التوزيعات الاحتمالية وهي التوزيعات المنفصلة.

1.3.5 التوزيعات الاحتمالية المنفصلة (Discrete Probability Distributions)

تعريف (2.5): **التوزيع الاحتمالي المنفصل** (Discrete Probability Distribution): إذا كان X متغير عشوائي يأخذ القيم المنفصلة $x_1, x_2, \dots, x_n$ باحتمالات مناظرة $P(x_1), P(x_2), \dots, P(x_n)$ فإن الدالة $f(x) = P(X = x)$ تعرّف بأنها دالة احتمالية أو توزيع احتمالي منفصل إذا كان:

1. $0 \leq P(X = x) \leq 1 \ , \forall x$
2. $\sum_x P(X = x) = 1$

وسنقوم باستخدام رمز الدالة $P(X = x)$ أو $P(x)$ أو $P_X(x)$ للتعبير عن التوزيع الاحتمالي المنفصل ضمن أجزاء هذا الكتاب بحسب الحاجة، ويطلق عادة على الدالة $P(x)$ دالة الكتلة الاحتمالية (Probability Mass Function).

من التعريف (2.5) يتضح لنا أن الدالة الرياضية تسمى بالدالة الاحتمالية أو التوزيع الاحتمالي إذا ما كانت كل قيمها موجبة ومحصورة بين القيمتين 0 و 1 ، (وهذا من مسلمات الاحتمال)، وإذا ما كان مجموع الاحتمالات المناظرة للمتغير العشوائي يساوي الواحد الصحيح.

ملاحظات:

1. الشكل العام للتوزيع الاحتمالي المنفصل يكون عادة بصورة الجدول التالي:

جدول (2.5): الشكل العام للتوزيع الاحتمالي المنفصل.

x	x_1	x_2	$\cdots$	x_n
$P(X = x)$	$P(x_1)$	$P(x_2)$	$\cdots$	$P(x_n)$

2. عندما يأخذ المتغير العشوائي X قيما منفصلة ولكن غير منتهية $x_1, x_2, \dots$ فإن الاحتمالات المناظرة تكون $P(x_1), P(x_2), \dots$ على التوالي.
3. الرمز $\sum_x$ يرمز إلى أن المجموع يتم أخذه على كل "الخلايا" في جدول التوزيع الاحتمالي.

مثال (2.5): في تجربة إلقاء عملتين معدنيتين مرة واحدة فإن فراغ العينة هو $S = \{HH, HT, TH, TT\}$، وإذا ما تم تعريف المتغير العشوائي X بأنه يمثل عدد الصور فإن قيمه ستكون كما هو موضح في جدول (3.5).

جدول (3.5): قيم المتغير العشوائي X واحتمالاتها المناظرة للمثال (2.5).

فراغ العينة	HH	HT	TH	TT
X	2	1	1	0
$P(X = x)$	$1/4$	$1/4$	$1/4$	$1/4$

ويمكن "اختزال" قيم المتغير العشوائي X في الجدول السابق (3.5) لتكوين توزيعه الاحتمالي بشكله التقليدي كما هو موضح في الجدول التالي (جدول (4.5)).

جدول (4.5): التوزيع الاحتمالي للمتغير العشوائي X للمثال (2.5).

X	0	1	2
$P(X=x)$	$1/4$	$1/2$	$1/4$

ولاحظ من جدول (3.5) أن القيمة $x = 2$ كان الاحتمال المناظر لها هو $1/4$ لأن احتمال الحصول على هذه القيمة ($P(X=2) =$ $1/4$) هو مرة واحدة ضمن الحالات الأربع لفراغ العينة، وهكذا بالنسبة لباقي قيم X. ولاحظ أيضا أن جدول التوزيع الاحتمالي (جدول (4.5)) يتم فيه كتابة قيم X عادة بالترتيب التصاعدي وبدون كتابة قيم X المكررة، ويؤخذ ذلك بالاعتبار عند كتابة الاحتمالات المناظرة لها، وهذا ما نلاحظه في القيمة $x = 1$ حيث أنها تكررت مرتين، فنقوم بجمع الاحتمالات المناظرة لكل تكرار فنحصل على $1/4 + 1/4 = 1/2$. ويمكن ببساطة أن نثبت أن جدول (4.5) يمثل توزيع احتمالي لأن كل احتمال فيه هو موجب وأن مجموع الاحتمالات يساوي الواحد الصحيح؛

$$\sum_{i=1}^{3} P(x_i) = 1/4 + 1/2 + 1/4 = 1$$

مثال (3.5): في قسم الأمراض السارية في أحد المستشفيات يوجد 12 مريضا منهم 3 مرضى مصابين بإنفلونزا الخنازير. فإذا ما تم اختيار 3 مرضى (عينة) من هذا القسم بصورة عشوائية، أوجد:

1. التوزيع الاحتمالي لعدد مرضى إنفلونزا الخنازير.
2. احتمال عدم وجود أي مريض بالأنفلونزا ضمن العينة المسحوبة.
3. احتمال وجود من مريض بالأنفلونزا إلى ثلاثة مرضى ضمن العينة.
4. احتمال وجود مريضين بالأنفلونزا على الأقل ضمن العينة.
5. احتمال وجود مريض واحد بالأنفلونزا على الأكثر ضمن العينة.

الحل:

1. لدينا العدد الكلي للمرضى $n = 12$ مريض ونريد اختيار عينة مكونة من $r = 3$ منهم باعتبار وجود 3 مرضى لهم الصفة المطلوبة (وهي الإصابة بمرض إنفلونزا الخنازير) للمتغير العشوائي، وبالتالي فإن قيم المتغير ستكون $x = 0, 1, 2, 3$ حيث تشير القيمة $x = 0$ إلى عدم وجود الصفة، والقيمة $x = 1$ إلى وجود مريض واحد له الصفة، ... وهكذا. نقوم أولا بحساب الاحتمالات المناظرة لكل قيمة من قيم X كالتالي[1]:

$$P(X=x) = \frac{C_x^k C_{r-x}^{n-k}}{C_r^n} \quad , x = 0, 1, 2, 3$$

حيث k هي عدد المرضى الذين يتمتعون بالصفة وعددهم 3، فيكون؛

$$P(X=0) = \frac{C_0^3 C_3^9}{C_3^{12}} = \frac{84}{220} = 0.38$$

[1] يعرف هذا التوزيع الاحتمالي بالتوزيع **فوق الهندسي** والذي سيتم التطرق إليه في الفصل السادس.

$$P(X=1)=\frac{C_1^3 C_2^9}{C_3^{12}}=\frac{108}{220}=0.49$$

$$P(X=2)=\frac{C_2^3 C_1^9}{C_3^{12}}=\frac{27}{220}=0.12$$

$$P(X=3)=\frac{C_3^3 C_0^9}{C_3^{12}}=\frac{1}{220}=0.01$$

ويكون التوزيع الاحتمالي للمتغير X بالصورة:

جدول (5.5): التوزيع الاحتمالي لعدد مرضى إنفلونزا الخنازير (مثال (3.5))

x	0	1	2	3
$P(x)$	0.38	0.49	0.12	0.01

2. احتمال عدم وجود أي مريض بإنفلونزا الخنازير ضمن العينة المسحوبة يعني أن يأخذ المتغير العشوائي القيمة 0 هو $P(X=0)=P(0)=0.38$.
3. احتمال وجود من مريض بالأنفلونزا إلى ثلاثة مرضى هو

$$P(1\le X\le 3)=P(X=1)+P(X=2)+P(X=3)$$
$$=0.49+0.12+0.01=0.62$$

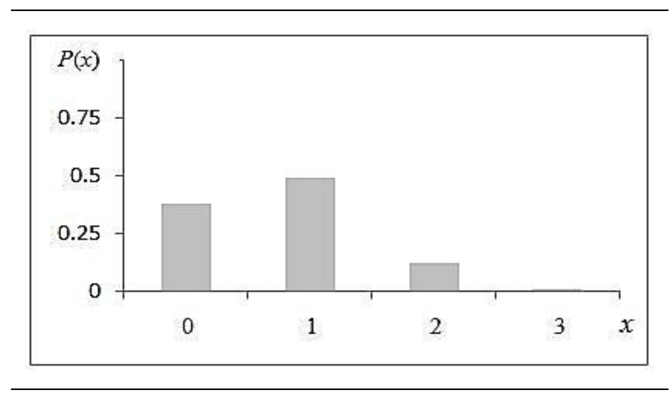
4. احتمال وجود مريضين بالأنفلونزا على الأقل ضمن العينة هو

$$P(X\ge 2)=P(X=2)+P(X=3)=0.12+0.01=0.13$$

5. احتمال وجود مريض واحد بالأنفلونزا على الأكثر ضمن العينة يعني احتمال وجود مريض واحد بالأنفلونزا فأقل، أي

$$P(X\le 1)=P(X=0)+P(X=1)=0.38+0.49=0.87$$

يمكن ملاحظة أن جدول التوزيع الاحتمالي يشبه إلى حد كبير جدول التوزيع التكراري إذا ما اعتبرنا أن قيم المتغير العشوائي X تعبر عن قيم مفردات البيانات أو مراكز الفترات وقيم الاحتمالات تعبر عن قيم التكرارات مقسومة على عدد المفردات الكلية. وبناء على هذا التشابه يمكننا استخدام الرسم البياني لتمثيل جداول التوزيع الاحتمالي كما هو الحال مع التوزيعات التكرارية. ومن أبرز هذه الرسومات البيانية هو **المدرج الاحتمالي** (Probability Histogram)، والذي يمكن استخدام جدول (5.5) في المثال (3.5) لرسمه. والشكل (1.5) يوضح المدرج الاحتمالي. ويمكن أيضا حساب الاحتمالات بصورة تصاعدية (تراكمية) كما هو الحال عند حساب التكرار المتجمع الصاعد من جداول التوزيع التكراري.



شكل (1.5): المدرج الاحتمالي لقيم المتغير X في المثال (3.5).

تعريف (3.5): **دالة التوزيع التراكمية (التجميعية)** (Cumulative Distribution Function): إذا كان X متغير عشوائي له دالة الكتلة الاحتمالية $P(x)$ فإن الدالة F والتي تحقق قيمها لكل عدد حقيقي x العلاقة $F(x) = P(X \leq x)$ تعرف[1] بدالة التوزيع التراكمية للمتغير العشوائي X.

مثال (4.5): من المثال السابق (3.5) نستطيع حساب الاحتمالات التراكمية عند أي نقطة x ، فمثلا:

$$F(2) = P(X \leq 2) = P(0) + P(1) + P(2) = 0.38 + 0.49 + 0.12 = 0.99$$

ولاحظ أن قيمة دالة التوزيع التراكمية لأدنى قيمة في جدول التوزيع الاحتمالي تساوي احتمال تلك القيمة؛ $F(min(x)) = P(min(x))$ وقيمة دالة التوزيع التراكمية لأعلى قيمة في الجدول تساوي الواحد الصحيح لأن ذلك يشمل مجموع كل الاحتمالات المناظرة لقيم X في جدول التوزيع الاحتمالي؛ $F(max(x)) = 1$ ، فمن المثال (3.5) نجد أن

$$F(0) = P(0) = 0.38$$

$$F(3) = P(X \leq 3) = P(0) + P(1) + P(2) + P(3) = 1$$

تعريف (4.5): **خواص دالة التوزيع التراكمية:** إذا كان X متغير عشوائي له دالة التوزيع التراكمية $F(x)$ فإن:

1. يوجد عدد حقيقي (m مثلا) بحيث تكون $F(x) = 0$ إذا كان $x < m$ ، ويوجد عدد حقيقي (M مثلا) بحيث تكون $F(x) = 1$ إذا كان $x \geq M$.
2. $F(x)$ هي دالة غير تناقصية، أي أن $F(x_1) \geq F(x_2)$ إذا كان $x_1 \geq x_2$.
3. $F(x)$ هي دالة سُلّمية (Step Function) بعدد منتهي من الدرجات.
4. لأي قيمة x لدينا $P(X > x) = 1 - F(x)$.
5. لأي قيمتين x_1 و x_2 بحيث $x_1 < x_2$ فإن
$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$$

وننتقل الآن إلى تعريف النوع الثاني من التوزيعات الاحتمالية وهي التوزيعات المتصلة.

2.3.5 التوزيعات الاحتمالية المتصلة (Continuous Probability Distributions)

تعريف (5.5): **التوزيع الاحتمالي المتصل أو المستمر** (Continuous Probability Distribution) : إذا كان X متغير عشوائي متصل (يأخذ قيم غير منفصلة ضمن فترة) فإن دالته الاحتمالية[2] (والتي يرمز لها بالرمز $f(x)$ أو $f_X(x)$) تكون دالة تتوفر فيها الشروط التالية:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$ ، حيث $(-\infty ، \infty)$ هي الفترة الافتراضية التي ينتمي لها X.

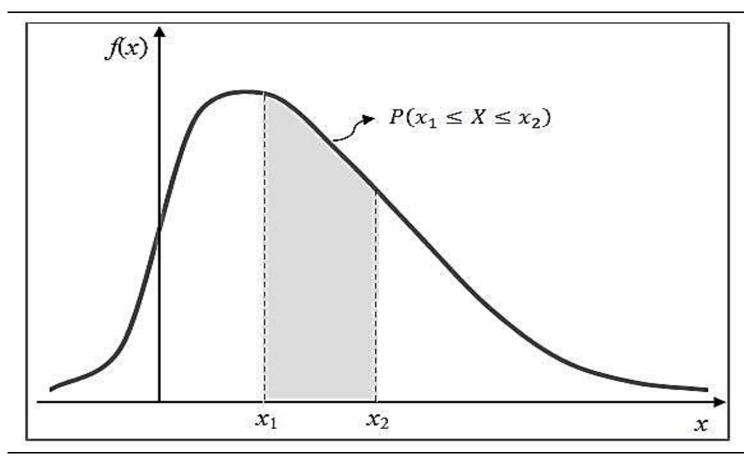
[1] بعض الكُتّاب يطلقون على دالة التوزيع التراكمية مصطلح "**دالة التوزيع**" أو "**دالة الاحتمال**".

[2] والتي تعرف أيضا بدالة الكثافة الاحتمالية (Probability Density Function).

3. المساحة تحت منحنى الدالة $f(x)$ من القيمة x_1 إلى القيمة x_2 ، والتي تساوي احتمال أن تقع قيمة المتغير X بين x_1 و x_2 ، يتم حسابها بالصورة:

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)dx$$

ويمكن تمثيل هذه المساحة أو هذا الاحتمال كما هو موضح في الشكل (2.5).



شكل (2.5): تمثيل الاحتمال $P(x_1 \leq X \leq x_2)$ تحت منحنى الدالة $f(x)$.

ملاحظات:

1. إن مفهوم دالة التوزيع التراكمية $F(x)$ هو نفسه في التوزيعات المتصلة مع ملاحظة أن تراكم أو تجمع قيم المتغير العشوائي X لا يكون بجمع القيم، كما هو الحال مع التوزيعات المنفصلة، بل بأخذ التكامل، فمثلا إذا كان من المطلوب حساب قيمة F عند النقطة $x = a$ فإن:

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

2. يمكننا كتابة $f(x) = \frac{\partial F(x)}{\partial x}$ ، بمعنى أنه يمكن الوصول لدالة الكثافة الاحتمالية $f(x)$ عن طريق أخذ التفاضل الأول لدالة التوزيع التراكمية لأن الأخيرة تمثل تكامل دالة الكثافة الاحتمالية.

3. لأي قيمة $x \in (-\infty, \infty)$ يكون $P(X > x) = 1 - P(X \leq x)$.

مثال (5.5): إذا كان X متغير عشوائي متصل له الدالة التالية:

$$f(x) = \begin{cases} \frac{x}{2} \ ; 0 \leq x \leq 2 \\ 0 \ ; \ \text{خلاف ذلك} \end{cases}$$

1. أثبت أن $f(x)$ هي دالة كثافة احتمالية.
2. أوجد $P(X > 1)$.
3. أوجد $P(0 \leq X \leq 1)$.
4. أوجد $F(1)$ و $F(2)$.

الحل:

1. لإثبات أن $f(x)$ هي دالة كثافة احتمالية يجب إثبات أن تكاملها على نطاق الفترة المعرفة عليها يساوي الواحد الصحيح وأن الدالة موجبة؛

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{2} \frac{x}{2}dx = \frac{1}{2}\left[\frac{x^2}{2}\right]_0^2 = \frac{1}{2}\left[\frac{4}{2} - 0\right] = 1$$

وحيث أن $f(x) = \frac{x}{2}$ هي دالة موجبة، نستطيع القول بأنها دالة كثافة احتمالية.

2.

$$P(X > 1) = P(1 < X \leq 2) = \int_{1}^{2} \frac{x}{2}dx = \frac{1}{2}\left[\frac{x^2}{2}\right]_1^2 = \frac{1}{2}\left[\frac{4}{2} - \frac{1}{2}\right] = \frac{3}{4}$$

3.

$$P(0 \leq X \leq 1) = \int_{0}^{1} \frac{x}{2}dx = \frac{1}{2}\left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}\left[\frac{1}{2} - 0\right] = \frac{1}{4}$$

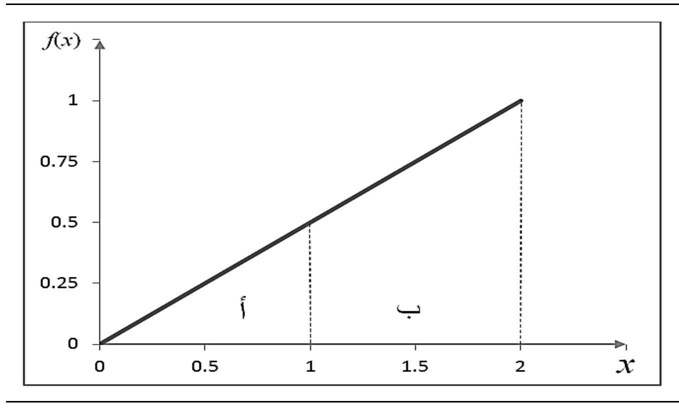
4.

$$F(1) = P(X \leq 1) = \int_{-\infty}^{1} f(x)dx = \int_{0}^{1} \frac{x}{2}dx = \frac{1}{4}$$

وكذلك

$$F(2) = F(\max(x)) = \int_{0}^{2} \frac{x}{2}dx = 1$$

وللمزيد من التوضيح يمكننا رسم شكل الدالة $f(x) = \frac{x}{2}$ وتعيين الاحتمالات المحسوبة في المثال بيانيا كما هو في شكل (3.5).



شكل (3.5): شكل الدالة $f(x) = \frac{x}{2}$ في المثال (5.5).

حيث تُمثل:

- المساحة الكلية للمثلث:

$\int_0^2 \frac{x}{2}dx = P(0 \leq X \leq 2) = 1$

- المساحة (أ):

$P(0 \leq X \leq 1) = \frac{1}{4}$

- المساحة (ب):

$P(X > 1) = \frac{3}{4}$

مثال (6.5): إذا كان X متغير عشوائي متصل له دالة الكثافة الاحتمالية التالية:

$$f(x) = \begin{cases} \frac{1}{6}x + c\,; & 0 \le x \le 3 \\ 0\,; & \text{خلاف ذلك} \end{cases}$$

فأوجد قيمة الثابت c.

الحل:

حيث أن الدالة $f(x)$ هي دالة كثافة احتمالية (كما هو معطى في المثال) فإنها تتمتع بالخاصية:

$$\int_0^3 f(x)dx = 1$$

وبالتالي

$$\int_0^3 \left(\frac{1}{6}x + c\right)dx = 1$$

$$\frac{1}{6}\left[\frac{x^2}{2}\right]_0^3 + c[x]_0^3 = 1$$

$$\frac{1}{6}\left[\frac{9}{2} - 0\right] + c[3 - 0] = 1$$

$$\frac{9}{12} + 3c = 1$$

$$3c = \frac{3}{12}$$

$$c = \frac{1}{12}$$

فتكون

$$f(x) = \frac{1}{6}x + \frac{1}{12} \quad , \quad 0 \le x \le 3$$

4.5 التوقع والتباين للمتغيرات العشوائية (Expectation and Variance for Random Variables)

كما هو الحال عند حساب الوسط الحسابي والتباين لأي مجموعة من البيانات، فإنه يمكن حساب هذين المقياسين لوصف أو تلخيص التوزيعات الاحتمالية لأي متغير عشوائي، حيث يصف الوسط الحسابي (أو التوقع (Expectation) كما يطلق عليه عادة في نظرية الاحتمال) مركز التوزيع الاحتمالي، ويعبر التباين أو الانحراف المعياري عن مدى تشتت توزيع المتغير العشوائي. وبالإضافة لذلك، فإن التوقع والتباين تستخدم كمقاييس للمقارنة بين التوزيعات الاحتمالية المختلفة.

ومفهوم هذين المقياسين هو واحد بالنسبة للتوزيعات الاحتمالية المنفصلة والمتصلة، إلا أن الصيغة الرياضية المستخدمة في الحساب تختلف باختلاف طبيعة المتغير العشوائي.

1.4.5 التوقع والتباين للتوزيع الاحتمالي المنفصل
(Expectation and Variance for Discrete Distribution)

تعريف (6.5): **توقع المتغير العشوائي المنفصل:** إذا كان X متغير عشوائي منفصل يأخذ القيم $x_1, x_2, \ldots, x_n$ وله دالة الكتلة الاحتمالية $P(x_i)$ حيث $i = 1, 2, \ldots, n$ ، فإن التوقع للمتغير X (ويرمز له بالرمز $E(X)$) يعرّف بالصورة:

$$E(X) = \sum_{i=1}^{n} x_i\, P(x_i) = x_1 P(x_1) + \cdots + x_n P(x_n)$$

وكتعميم للتعريف السابق، يمكن أيضا حساب التوقع لأي دالة في المتغير العشوائي كما يوضح التعريف التالي:

تعريف (7.5): **توقع دالة في المتغير العشوائي المنفصل:** إذا كان X متغير عشوائي منفصل يأخذ القيم $x_1, x_2, \ldots, x_n$ وله دالة الكتلة الاحتمالية $P(x_i)$ حيث $i = 1, 2, \ldots, n$ ، وكانت $g(x)$ هي دالة في المتغير X ، فإنه يمكن حساب توقع $g(x)$ بالصورة:

$$E\big(g(x)\big) = \sum_{i=1}^{n} g(x_i) P(x_i)$$

إلا أنه لحساب توقع $g(x)$ بصورة عملية، يتم حساب توقع المتغير X أولا ثم استخدام خواصه لحساب توقع $g(x)$ كما سنرى في النظرية التالية:

نظرية (1.5): خواص توقع المتغير العشوائي[1]:

إذا كان X متغير عشوائي وكان a و b هما أي عددين (ثابتين) فإن

$$E(aX \pm b) = aE(X) \pm b$$

الإثبات:

بافتراض أن المتغير X يأخذ قيما منفصلة $x_1, x_2, \ldots, x_n$ يكون لدينا

$$\begin{aligned} E(aX + b) &= \sum_{i=1}^{n} (ax_i + b)P(x_i) = (ax_1 + b)P(x_1) + \cdots + (ax_n + b)P(x_n) \\ &= a[x_1P(x_1) + \cdots + x_nP(x_n)] + b[P(x_1) + \cdots + P(x_n)] \\ &= a\sum_{i=1}^{n} x_iP(x_i) + b\sum_{i=1}^{n} P(x_i) = aE(X) + b \end{aligned}$$

حيث $\sum_{i=1}^{n} P(x_i) = 1$.

---

[1] تنطبق هذه الخواص على المتغير العشوائي المتصل أيضا.

ويمكن، من التعريف السابق ملاحظة أن $E(a) = a$ ، بمعنى أن توقع الثابت يساوي الثابت نفسه.

مثال (7.5): في تجربة إلقاء عملة معدنية ثلاث مرات تم تعريف المتغير العشوائي X بأنه يمثل عدد مرات ظهور الصورة (H). أوجد:

1. $E(X)$ 2. $E(3X - \frac{1}{4})$ 3. $E\big(g(x)\big)$ إذا كانت $g(x) = 2x + 1$

الحل:

التوزيع الاحتمالي للمتغير X من هذه التجربة العشوائية هو

x	0	1	2	3
$P(x)$	1/8	3/8	3/8	1/8

1. لدينا

$$E(X) = \sum_{i=1}^{n} x_i P(x_i) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{3}{2} = 1.5$$

ولاحظ أنه ليس بالضرورة أن تكون قيمة التوقع هي إحدى القيم المشاهدة للمتغير العشوائي X ، وأنه في هذا المثال لا معنى منطقي لأن نقول أن عدد الصور المتوقع (أو الوسط الحسابي لعدد الصور) هو 1.5 صورة، ولكن كما وضحنا سابقا أن حساب قيمة التوقع في التوزيع الاحتمالي يكون للدلالة على مركز التوزيع، ففي مثالنا هذا نستطيع القول أن عدد الصور في الرميات الثلاث يتركز مابين صورة واحدة وصورتين، وهذا منطقي لأن حاصل جمع الاحتمالين المناظرين للقيم $x = 1, 2$ وهو $3/8 + 3/8 = 6/8$ يمثل القيمة الكبرى للاحتمال.

2. من خواص التوقع

$$E(3X\text{-}1/4) = 3E(X) - \frac{1}{4} = 3 \times \frac{3}{2} - \frac{1}{4} = \frac{17}{4} = 4.25$$

3. حيث أن $E(X) = 3/2$ و $g(x) = 2X + 1$ فإن

$$E\big(g(x)\big) = E(2X + 1) = 2E(X) + 1 = 2 \times \frac{3}{2} + 1 = 4$$

تعريف (8.5): **تباين المتغير العشوائي المنفصل:** إذا كان X متغير عشوائي منفصل يأخذ القيم $x_1, x_2, \dots, x_n$ وله دالة الكتلة الاحتمالية $P(x_i)$ حيث $i = 1, 2, \dots, n$ ، فإن تباين X (والذي يرمز له بالرمز $Var(X)$)، يعرّف بالصورة:

$$Var(X) = E(X - E(X))^2$$

أو

$$= \sum_{i=1}^{n} \big(x_i - E(X)\big)^2 P(x_i)$$

أو

$$= \sum_{i=1}^{n} x_i^2 P(x_i) - \Big[\sum_{i=1}^{n} x_i P(x_i)\Big]^2$$

أو

$$= E(X^2) - [E(X)]^2$$

نظرية (2.5): خواص تباين المتغير العشوائي:

إذا كان X متغير عشوائي (منفصل أو متصل)، وكان a و b هما أي عددين (ثابتين) فإن

$$Var(aX \pm b) = a^2 Var(X)$$

الإثبات[1]:

$$Var(X) = E[(aX+b) - E(aX+b)]^2 = E[aX + b - [aE(X) + b]]^2$$
$$= E[aX + b - aE(X) - b]^2 = E[a(X - E(X)]^2$$
$$= E[a^2(X - E(X))]^2 = a^2 E(X - E(X))^2 = a^2 Var(X)$$

ولاحظ من النظرية السابقة أن $Var(a) = 0$ ، بمعنى أن تباين القيمة الثابتة يساوي صفرا.

مثال (8.5): من المثال (7.5) أوجد:

1. $Var(X)$ 2. $Var(3X-2)$ 3. $Var(Y)$ ، حيث $Y = \frac{X}{2} + 1$

الحل:

1. لدينا

$$E(X) = \frac{3}{2}$$

$$E(X^2) = \sum_{i=1}^{n} x_i^n P(x_i) = 0^2 \times \frac{1}{8} + 1^2 \times \frac{3}{8} + 2^2 \times \frac{3}{8} + 3^2 \times \frac{1}{8} = 3$$

$$Var(X) = E(X^2) - [E(X)]^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}$$

.2

$$Var(3X-2) = 9Var(X) = 9 \times \frac{3}{4} = \frac{27}{4}$$

.3

$$Var(Y) = Var\left(\frac{X}{2} + 1\right) = \frac{1}{4}Var(X) = \frac{1}{4} \times \frac{3}{4} = \frac{3}{16}$$

ولتوضيح مفهوم التباين أو التشتت في قيم المتغير العشوائي نأخذ المثال التالي:

مثال (9.5): إذا كان لدينا التوزيعين الاحتماليين التاليين (جدول (6.5))، للمتغيرين العشوائيين X_1 و X_2، فقارن بين درجتي تشتت كل من التوزيعين.

جدول (6.5): التوزيع الاحتمالي للمتغير X_1 و X_2 (مثال (9.5)).

توزيع X_2				
x	2	4	6	8
$P_2(x)$	1/8	2/8	3/8	2/8

توزيع X_1				
x	1	2	3	4
$P_1(x)$	1/8	2/8	3/8	2/8

[1] سيتم إثبات الخاصية في حالة الجمع، علما بأن الاثبات مماثل في حالة الطرح.

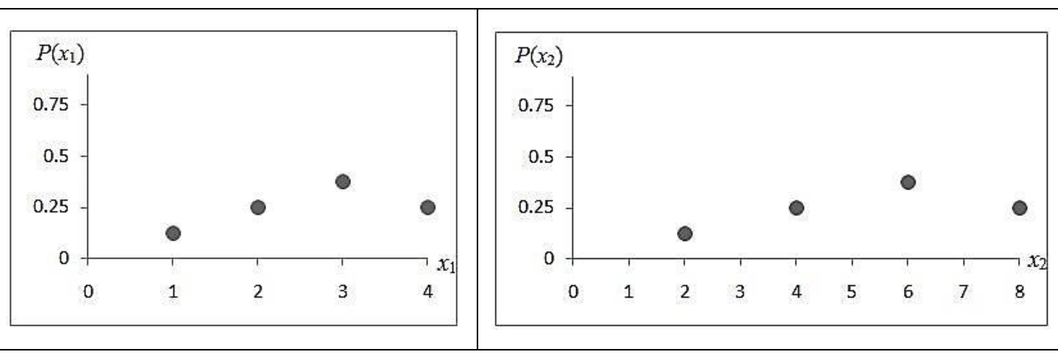
الحل:

لدينا من التوزيعان:

$$E(X_2) = \frac{11}{2}$$
$$E(X_2^2) = 34$$
$$Var(X_2) = 3.75$$
$$S.D(X_2) = \sqrt{Var(X_2)}$$
$$= \sqrt{3.75} = 1.94 \cong 2$$

$$E(X_1) = \frac{22}{8}$$
$$E(X_1^2) = \frac{68}{8}$$
$$Var(X_1) = 0.94$$
$$S.D(X_1) = \sqrt{Var(X_1)}$$
$$= \sqrt{0.94} = 0.97 \cong 1$$

وباستخدام الرسم البياني لتمثيل كلا من التوزيعين على شكل انتشار احتمالي نقطي، كما هو موضح في الشكل (4.5)، نحصل على صورة واضحة لتشتت التوزيعين.



شكل (4.5): شكل الانتشار الاحتمالي النقطي للتوزيعين الاحتماليين في المثال (9.5).

يلاحظ من الشكل السابق أنه رغم تمتع كلا من المتغيرين بنفس قيم الاحتمالات المناظرة وبنفس الترتيب، إلا أنهما يحظيان بدرجتي تشتت مختلفتين، حيث أن درجة تشتت توزيع المتغير X_2 هو ضعف درجة تشتت توزيع المتغير X_1 تقريبا.

2.4.5 التوقع والتباين للتوزيع الاحتمالي المتصل

(Expectation and Variance for Continuous Distribution)

تعريف (9.4): **توقع المتغير العشوائي المتصل:** إذا كان X متغير عشوائي متصل ومعرّف على الفترة $[a , b]$ ، وكانت له دالة الكثافة الاحتمالية $f(x)$ ، فإن توقع X يعرّف بالصورة:

$$E(X) = \int_a^b x\, f(x)dx$$

تعريف (10.4): **تباين المتغير العشوائي المتصل:** إذا كان X متغير عشوائي متصل ومعرّف على الفترة $[a, b]$ وكانت له دالة الكثافة الاحتمالية $f(x)$ ، فإن تباين المتغير X يعرّف بالصورة:

$$Var(X) = E(X - E(X))^2$$

أو

$$= \int_a^b \left(x - E(x)\right)^2 f(x)dx$$

أو

$$\int_a^b x^2 f(x)dx - \left[\int_a^b x f(x)dx\right]^2$$

أو

$$= E(X^2) - [E(X)]^2$$

مثال (10.5): مستخدما دالة الكثافة الاحتمالية في المثال (5.5) أوجد:

1. $E(X)$ 2. $E(5X)$ 3. $Var(X)$ 4. $S.D(X)$ 5. $Var(2X-1)$

الحل:

.1

$$E(X) = \int_0^2 xf(x)dx = \frac{1}{2}\int_0^2 xxdx = \frac{1}{2}\left[\frac{x^3}{3}\right]_0^2 = \frac{1}{2}\left[\frac{8}{3} - 0\right] = \frac{4}{3}$$

.2

$$E(5X) = 5E(X) = 5 \times \frac{4}{3} = \frac{20}{3}$$

.3

$$E(X^2) = \int_0^2 x^2 f(x)dx = \frac{1}{2}\int_0^2 x^2 xdx = \frac{1}{2}\left[\frac{x^4}{4}\right]_0^2 = \frac{1}{2}\left[\frac{16}{4} - 0\right] = 2$$

$$E(X^2) - [E(X)]^2 = 2 - (\frac{4}{3})^2 = \frac{2}{9}$$

.4

$$S.D(X) = \sqrt{Var(X)} = \sqrt{\frac{2}{9}} = \frac{\sqrt{2}}{3}$$

.5

$$Var(2X-1) = 4Var(X) = 4 \times \frac{2}{9} = \frac{8}{9}$$

في الفصل الثاني، تم عرض نظرية تشيبيشف التي تتعلق بتشتت قيم المفردات حول المركز، وفي هذا البند سيتم عرض متباينة تشيبيشف التي تمثل قاعدة هامة تتعلق بانتشار قيم المتغير العشوائي حول قيمة توقعه.

نظرية (3.5): متباينة تشيبيشف (Chebyshev's Inequality)

إذا كان X متغير عشوائي له التوقع $E(X)$ والتباين $Var(X)$ ، وكان $c > 0$ هو أي عدد، فإن:

$$P(|X - E(X)| > c) < \frac{Var(X)}{c^2}$$

بمعنى أن احتمال قيمة X ، والتي تبعد عن الوسط بمقدار أكبر من c، لن يتعدى القيمة $\frac{Var(X)}{c^2}$.

مثال (11.5): إذا كان X متغير عشوائي له دالة الكثافة الاحتمالية:

$$f(x) = \frac{1}{2\sqrt{3}}\,, \qquad -\sqrt{3} < x < \sqrt{3}$$

وكان $E(X) = 0$ ، $Var(X) = 1$ ، و $c = \frac{3}{2}$ فإن:

$$P(|X - E(X)| > c) = P\left(|X| \geq \frac{3}{2}\right) = 1 - P\left(|X| < \frac{3}{2}\right) = 1 - P\left(-\frac{3}{2} < X < \frac{3}{2}\right)$$

$$= 1 - \int_{-\frac{3}{2}}^{\frac{3}{2}} \frac{1}{2\sqrt{3}}\,dx = 1 - \frac{1}{2\sqrt{3}}[x]_{-3/2}^{3/2} = 0.13$$

وحيث أن

$$\frac{Var(X)}{c^2} = \frac{1}{(\frac{3}{2})^2} = \frac{4}{9} = 0.44$$

يكون

$$P(|X - E(X)| > c) < \frac{Var(X)}{c^2}$$

5.5 التوزيعات الاحتمالية المشتركة (Joint Probability Distributions)

لاحظنا فيما سبق كيف أنه يمكن تعريف أكثر من متغير عشوائي مختلف على نفس التجربة العشوائية أو نفس فراغ العينة، ومن ثمة إيجاد التوزيع الاحتمالي لكل منها على حدى. ولكن في كثير من الأحيان يكون من المطلوب دراسة أو إيجاد احتمال تقاطع (تداخل) قيم متغيرين أو أكثر، بغرض إيجاد توزيع احتمالي مشترك تجتمع فيه قيم أكثر من متغير عشوائي واحد بحيث تكون لها احتمالات مشتركة.

عندئذ سيتم التعامل مع توزيعات احتمالية (منفصلة أو متصلة) مشتركة تضم عدة متغيرات لها احتمالات مناظرة مشتركة فيما بينها، مثل إيجاد التوزيع الاحتمالي المشترك للطول والوزن لعينة من الأشخاص، أو التوزيع الاحتمالي المشترك لدرجات طلبة في ثلاث كليات مختلفة ... وغير ذلك من الأمثلة التي تعتمد على التعامل مع أكثر من متغير في آن واحد. وفي هذا الكتاب سيتم تناول الحالة المتعلقة بالتوزيعات الاحتمالية المشتركة لمتغيرين فقط، والتي تعرف **بالتوزيعات الثنائية** (Bivariate Distributions)، وسنبدأ بالتوزيعات المشتركة المنفصلة.

1.5.5 التوزيعات الاحتمالية المشتركة المنفصلة (الثنائية)(Joint Discrete Probability Distributions)

وهي التوزيعات التي تنتج عن حساب الاحتمال المشترك (الثنائي) لمتغيرين عشوائيين منفصلين.

تعريف (11.5): **دالة الكتلة الاحتمالية المشتركة لمتغيران منفصلان:** إذا كان X_1 و X_2 متغيران عشوائيان منفصلان فإن دالة الكتلة الاحتمالية المشتركة أو الثنائية لهما تعرّف بالصورة:

$$P_{X_1,X_2}(x_1,x_2) = P(x_1,x_2) = P(X_1 = x_1, X_2 = x_2) \quad , \quad \forall x_1, x_2$$

وتحقق الشرطين التاليين:

1. $P(x_1,x_2) \geq 0 \quad , \quad \forall x_1, x_2$
2. $\sum_{x_1,x_2} P(x_1,x_2) = 1$

حيث ترمز القيمتين x_1, x_2 لكل القيم التي يأخذها المتغير العشوائي X_1 (ولتكن n_1 قيمة) والمتغير العشوائي X_2 (ولتكن n_2 قيمة) على الترتيب.

وكما هو الحال مع التوزيعات الاحتمالية الأحادية (لمتغير واحد)، يمكن إيجاد دالة التوزيع الاحتمالي التراكمية للتوزيع المشترك كما يوضح التعريف التالي:

تعريف (12.5): **دالة التوزيع التراكمي المشتركة لمتغيران منفصلان:** إذا كان X_1 و X_2 متغيران عشوائيان منفصلان لهما دالة الكتلة الاحتمالية المشتركة $P(x_1, x_2)$ فإن دالة التوزيع الاحتمالي التراكمية، والتي يرمز لها بالرمز $F(x_1, x_2)$، تعرّف بالصورة:

$$F(x_1,x_2) = P(X_1 \leq x_1, X_2 \leq x_2) = \sum_{t_1=-\infty}^{x_1} \sum_{t_2=-\infty}^{x_2} P(t_1,t_2)$$

حيث t_1, t_2 تشير للتغير في قيم المتغيرين X_1, X_2 حتى الوصول للقيم x_1, x_2 على الترتيب.

وتتمتع دالة التوزيع التراكمية F بالخواص التي نسردها من خلال النظرية التالية:

نظرية (4.5): إذا كان X_1 و X_2 متغيران عشوائيان منفصلان[1] وكان لهما دالة التوزيع التراكمية المشتركة $F(x_1, x_2)$ فإن:

1. $F(-\infty,-\infty) = F(-\infty,x_2) = F(x_1,-\infty) = 0$
2. $F(\infty,\infty) = 1$
3. إذا كانت x_1^* و x_2^* هي قيم يأخذها المتغيران X_1 و X_2 بحيث $x_1^* \geq x_1$ و $x_2^* \geq x_2$ فإن $F(x_1^*,x_2^*) - F(x_1^*,x_2) - F(x_1,x_2^*) + F(x_1,x_2) \geq 0$

---

[1] تنطبق هذه النظرية أيضا على المتغيرات العشوائية المتصلة.

ويمكن أيضا إيجاد (أو فصل) الدوال الاحتمالية الأحادية لكل متغير من المتغيرات التي تكوّن التوزيع الاحتمالي المشترك، والتي تسمى عندها **بالدوال الهامشية** (Marginal Function) كما يوضح التعريف التالي.

تعريف (13.5): **دالة الكتلة الاحتمالية الهامشية** (Marginal Probability Distribution): إذا كان X_1 و X_2 متغيران عشوائيان منفصلان لهما دالة الكتلة الاحتمالية المشتركة $P(x_1, x_2)$ فإن:

1. الدالة $P_1(x_1) = \sum_{x_2} P(x_1, x_2)$ تسمى بالدالة الهامشية للمتغير X_1 حيث
$$P_1(x_1) \geq 0 \ , \ \sum_{x_1} P_1(x_1) = 1$$

2. الدالة $P_2(x_2) = \sum_{x_1} P(x_1, x_2)$ تسمى بالدالة الهامشية للمتغير X_2 حيث
$$P_2(x_2) \geq 0 \ , \ \sum_{x_2} P_2(x_2) = 1$$

ويكون

$$\sum_{x_1} P_1(x_1) = \sum_{x_2} P_2(x_2) = \sum_{x_1} \sum_{x_2} P(x_1, x_2) = 1$$

مثال (12.5): صندوق يحتوي على ثلاث كرات حمراء وكرتان خضراوان. تم سحب عينة عشوائية مكونة من كرتين بدون إرجاع، وتم تعريف المتغيران X_1 و X_2 كالتالي؛

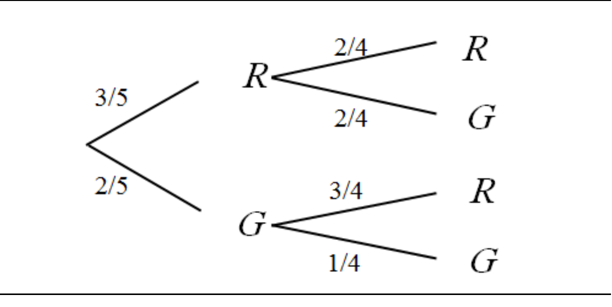
$$X_1 = \begin{cases} 0 & \text{إذا كانت الكرة الأولى المسحوبة خضراء} \\ 1 & \text{إذا كانت الكرة الأولى المسحوبة حمراء} \end{cases}$$

$$X_2 = \begin{cases} 0 & \text{إذا كانت الكرة الثانية المسحوبة خضراء} \\ 1 & \text{إذا كانت الكرة الثانية المسحوبة حمراء} \end{cases}$$

أوجد التوزيع الاحتمالي المشترك للمتغيرين X_1 و X_2 ، ثم أوجد دالة التوزيع الهامشية لكل منهما.

الحل:

لتبسيط إيجاد فراغ العينة لهذه التجربة العشوائية نستخدم الأشجار الاحتمالية، حيث سيُرمز للكرة الحمراء بالرمز R والكرة الخضراء بالرمز G فيكون:



شكل (5.5): الشجرة الاحتمالية لتجربة سحب الكرتين في المثال (12.5).

ويكون فراغ العينة

$$\begin{array}{cccccc} S & = & \{RR, & RG, & GR, & GG\} \\ & & \downarrow & \downarrow & \downarrow & \downarrow \\ X_1 & : & 1 & 1 & 0 & 0 \\ X_2 & : & 1 & 0 & 1 & 0 \end{array}$$

وبالتالي يمكن تكوين جدول التوزيع الاحتمالي المشترك، (جدول (7.5))، كما يلي:

جدول (7.5): التوزيع الاحتمالي المشترك لألوان الكرات المسحوبة في المثال (12.5).

X_1 \ X_2	0	1	المجموع
0	2/20	6/20	2/5
1	6/20	6/20	3/5
المجموع	2/5	3/5	1

حيث يكون الاحتمال $P(X_1= 0 , X_2= 0) = P(0,0)$ مثلا والذي يمثل احتمال الحصول على كرة أولى خضراء و كرة ثانية خضراء هو $P(0,0) = 2/5 \times 1/4 = 2/20$.

ومن العامود الأول (من اليمين) والصف الأخير في جدول (7.5) يمكن تكوين التوزيع الاحتمالي الهامشي للمتغيرين X_1 و X_2 على الترتيب كما هو موضح في جدول (8.5).

جدول (8.5): التوزيع الاحتمالي الهامشي للمتغير X_1 والمتغير X_2 في المثال (12.5).

X_1	0	1
$P_1(x_1)$	2/5	3/5

X_2	0	1
$P_2(x_2)$	2/5	3/5

1.1.5.5 الشرطية والاستقلال في التوزيعات المشتركة

(Conditionality and Independence in Joint Distributions)

تعريف (14.5): **التوزيع الاحتمالي الشرطي في التوزيعات المشتركة:** إذا كان X_1 و X_2 متغيران عشوائيان منفصلان لهما دالة الكتلة الاحتمالية المشتركة $P(x_1,x_2)$ فإن التوزيع الاحتمالي الشرطي للمتغير X_2 علما بأن $X_1= x_1$ يكون معرفا بالصورة:

$$P(X_2 = x_2 \backslash X_1 = x_1) = P(x_2 \backslash x_1) = \frac{P(x_1, x_2)}{P_1(x_1)} \quad , \quad P_1(x_1) > 0$$

حيث $P_1(x_1)$ هي الدالة الاحتمالية الهامشية للمتغير X_1 .

ملاحظة: يمكن بالمثل (من التعريف (14.5)) حساب دالة الاحتمال الشرطي للمتغير X_1 علما بأن $X_1= x_1$ بالصورة:

$$P(x_1 \backslash x_2) = \frac{P(x_1, x_2)}{P_2(x_2)} \quad , \quad P_2(x_2) > 0$$

مثال (13.5): من جدول التوزيع الاحتمالي المشترك التالي (جدول (9.5)) أوجد التوزيع الاحتمالي الشرطي للمتغير X_1 عندما $X_2 = 1$ ، ثم أوجد $P(X_1 = 0\backslash X_2 = 1)$.

جدول (9.5): التوزيع الاحتمالي المشترك للمتغيرين X_1 و X_2 في المثال (13.5).

X_1 \ X_2	0	1	2	المجموع
0	3/28	9/28	3/28	15/28
1	3/14	3/14	0	3/7
2	1/28	0	0	1/28
المجموع	5/14	15/28	3/28	1

الحل:

لدينا

$$P(x_1\backslash X_2 = 1) = P(x_1\backslash 1) = \frac{P(x_1, 1)}{P_1(1)} = \frac{P(x_1, 1)}{15/28} \quad , \quad x_1 = 0,1,2$$

وبالتعويض في الدالة السابقة عن قيم المتغير X_1 نحصل على

$$P(0\backslash 1) = \frac{P(0,1)}{15/28} = \frac{9/28}{15/28} = \frac{9}{15}$$

$$P(1\backslash 1) = \frac{P(1,1)}{15/28} = \frac{3/14}{15/28} = \frac{2}{5}$$

$$P(2\backslash 1) = \frac{P(2,1)}{15/28} = \frac{0}{15/28} = 0$$

وبالتالي فإن التوزيع الشرطي للمتغير X_1 عندما $X_2 = 1$ هو (جدول (10.5)):

جدول (10.5): التوزيع الاحتمالي الشرطي للمتغير X_1 عندما $X_2 = 1$ في المثال (13.5).

X_1	0	1	2
$P_1(x_1\backslash 1)$	9/15	2/5	0

ومن الجدول السابق يمكن حساب الاحتمال الشرطي

$$P(X_1 = 0\backslash X_2 = 1) = P(0\backslash 1) = \frac{9}{15}$$

تعريف (15.5): **الاستقلال في التوزيعات الاحتمالية المشتركة:** يقال أن المتغيران X_1 و X_2 متغيران مستقلان إذا وفقط إذا أمكن كتابة دالة التوزيع الاحتمالي المشتركة لهما كحاصل ضرب الدوال الهامشية لهما وذلك لكل قيم المتغيرين.

ففي حالة التوزيعات الاحتمالية المنفصلة إذا كان $P(x_1, x_2) = P_1(x_1) \times P_2(x_2) \quad , \quad \forall x_1, x_2$ فإن X_1 و X_2 يكونا متغيرين عشوائيين مستقلين.

وهذا يعني أنه إذا لم يتحقق التعريف السابق حتى ولو لنقطة (قيمة) واحدة من قيم المتغيرين فهذا ينفي وجود الاستقلال بينهما.

مثال (14.5): من المثال (12.5) نجد من جدول (7.5) أن

$$P(0,0) = \frac{2}{20} \neq P_1(0) \times P_2(0) = \frac{2}{5} \times \frac{2}{5} = \frac{4}{25}$$

وهذا يكفي للقول بأن المتغيرين X_1 و X_2 هما غير مستقلين.

2.1.5.5 التوقع والتباين للتوزيعات المشتركة (Expectation and Variance for Joint Distributions)

سنتناول فيما يلي التوقع والتباين وبعض خواصهما المتعلقة بالتوزيعات الاحتمالية، مع ملاحظة أن بعض الصيغ الرياضية والخواص قد تستخدم مع نوعي التوزيعات الاحتمالية؛ المنفصلة والمتصلة.

تعريف (16.5): **التوقع في التوزيعات الاحتمالية المشتركة المنفصلة:** إذا كان X_1 و X_2 متغيران عشوائيان منفصلان لهما دالة الكتلة الاحتمالية المشتركة $P(x_1,x_2)$ فإن الوسط الحسابي أو التوقع المشترك لهما يعرف بالصورة

$$E(X_1X_2) = \sum_{x_1}\sum_{x_2} x_1x_2\, P(x_1,x_2)$$

نظرية (5.5): التوقع المشترك للمتغيرين المستقلين (المنفصلين أو المتصلين) X_1 و X_2 هو حاصل ضرب توقع كل منهما في الآخر، بمعنى $E(X_1X_2) = E(X_1)E(X_2)$.

الإثبات[1]:

لدينا

$$E(X_1X_2) = \sum_{x_1}\sum_{x_2} x_1x_2P(x_1,x_2)$$

وحيث أن X_1 و X_2 هما مستقلان، فإن

$$P(x_1,x_2) = P_1(x_1) \times P_2(x_2)$$

وبالتالي

$$E(X_1X_2) = \sum_{x_1}\sum_{x_2} x_1x_2P(x_1)P(x_2) = \sum_{x_1} x_1P_1(x_1)\sum_{x_2} x_2P_2(x_2) = E(X_1)E(X_2)$$

مثال (15.5): من المثال (12.5) أوجد $E(X_1X_2)$ وكذلك $E(X_1)$ و $E(X_2)$.

الحل:

[1] في حالة المتغيرات المنفصلة.

من جدول (7.5) نجد أن:

$$E(X1X2) = 0 \times 0 \times \frac{2}{20} + 0 \times 1 \times \frac{6}{20} + 1 \times 0 \times \frac{6}{20} + 1 \times 1 \times \frac{6}{20} = \frac{6}{20}$$

و

$$E(X_1) = \sum_{x_1} x_1 P_1(x_1) = 0 \times \frac{2}{5} + 1 \times \frac{3}{5} = \frac{3}{5}$$

و

$$E(X_2) = \sum_{x_2} x_2 P_2(x_2) = 0 \times \frac{2}{5} + 1 \times \frac{3}{5} = \frac{3}{5}$$

ويلاحظ أن

$$E(X_1X_2) = \frac{6}{20} \neq E(X_1)E(X_2) = \frac{3}{5} \times \frac{3}{5} = \frac{9}{25}$$

وهذا يعني أن X_1 و X_2 هما غير مستقلين.

نظرية (6.5): (خواص التوقع في التوزيعات المشتركة)

إذا كان X_1 و X_2 متغيران عشوائيان (منفصلان أو متصلان) فإن:

1. $E(X_1 \pm X_2) = E(X_1) \pm E(X_2)$
2. $E(aX_1 \pm bX_2) = aE(X_1) \pm bE(X_2)$

الإثبات:

سيتم إثبات أن $E(X_1 + X_2) = E(X_1) + E(X_2)$ في حالة التوزيعات المنفصلة. لدينا

$$E(X_1 + X_2) = \sum_{x_1}\sum_{x_2}(x_1 + x_2)P(x_1, x_2) = \sum_{x_1}\sum_{x_2} x_1 P(x_1, x_2) + \sum_{x_1}\sum_{x_2} x_2 P(x_1, x_2)$$

$$= \sum_{x_1} x_1 \sum_{x_2} P_1(x_1, x_2)$$

$$+ \sum_{x_2} x_2 \sum_{x_1} P_1(x_1, x_2)$$

$$= \sum_{x_1} x_1 P_1(x_1) + \sum_{x_2} x_2 P_2(x_2) = E(X_1) + E(X_2)$$

ويمكن إثبات الفقرة (2) بنفس الطريقة مع مراعاة استخدام النظرية (1.5).

نظرية (7.5): (خواص التباين في التوزيعات المشتركة ذات المتغيرات المستقلة)

إذا كان X_1 و X_2 متغيران عشوائيان مستقلان (منفصلان أو متصلان) وكان $Var(X_1)$ و $Var(X_2)$ هما تبايني التوزيع الهامشي للمتغيرين على التوالي، فإن:

1. $Var(X_1 \pm X_1) = Var(X_1) + Var(X_2)$

2. $Var(aX_1 \pm bX_1) = a^2Var(X_1) + b^2Var(X_2)$

الإثبات:

سيتم إثبات الخاصية (1) في حالة الطرح للتوزيعات المنفصلة، علما بأن الخاصية (2) يمكن الوصول إليها بسهولة باستخدام خواص التباين العامة التي سبق تناولها.

من تعريف التباين نستطيع كتابة

$$\begin{aligned} Var(X_1 - X_2) &= E[(X_1 - X_2) - E(X_1 - X_2)]^2 = E[X_1 - X_2 - E(X_1) + E(X_2)]^2 \\ &= E[X_1 - E(X_1) - (X_2 - E(X_2))]^2 \\ &= E\big(X_1 - E(X_1)\big)^2 + E\big(X_2 - E(X_2)\big)^2 \\ &\quad - 2E\big[\big(X_1 - E(X_1)\big)\big(X_2 - E(X_2)\big)\big] \end{aligned}$$

وبتحليل المقدار الأخير

$$\begin{aligned} E\big[\big(X_1 - E(X_1)\big)\big(X_2 - E(X_2)\big)\big] &= E[X_1X_2 - X_2E(X_1) - X_1E(X_2) + E(X_1)E(X_2)] \\ &= E(X_1X_2) - E(X_1)E(X_2) - E(X_1)E(X_2) + E(X_1)E(X_2) \end{aligned}$$

وحيث أن X_1 و X_2 هما مستقلان، فإن $E(X_1X_2) = E(X_1)E(X_2)$ وبالتالي فإن المقدار الأخير سيساوي الصفر.

(ولاحظ أن $E(E(X_1)) = E(X_1)$ لأن توقع المتوسط (وهو ثابت) يساوي نفسه). وهكذا فإن

$$Var(X_1 - X_2) = Var(X_1) + Var(X_2)$$

مثال (16.5): إذا كان X_1 و X_2 متغيران عشوائيان مستقلان (منفصلان أو متصلان) لهما التوقعان $E(X_1) = 3$ و $E(X_2) = 4$ على التوالي، ولهما التباينان؛ $Var(X_1) = 1$ و $Var(X_2) = 2$ على التوالي فأوجد:

1. $E(5X_1 + 2X_2)$.
2. $Var(Z)$ حيث $Z = 3X_1 - 2X_2 + 5$.

الحل:

1. $E(5X_1 + 2X_2) = 5E(X_1) + 2E(X_2) = 5 \times 3 + 2 \times 4 = 23$

2. حيث أن المتغير Z يضم ثلاثة حدود، فبوضع $t_1 = 3X_1 - 2X_2$ و $t_2 = 5$ يكون المطلوب إيجاد التباين للفرق بين t_1 و t_2 كالتالي

$$Var(t_1 - t_2) = Var(t_1) + Var(t_2) = Var(3X_1 - 2X_2) + Var(5)$$
$$= Var(3X_1 - 2X_2) = 9Var(X_1) + 4Var(X_2)$$
$$= 9 \times 1 + 4 \times 2 = 17$$

3.1.5.5 التغاير والارتباط في التوزيعات المشتركة
(Covariance and Correlation in Joint Distribution)

من خلال دراستنا السابقة لمفهوم التباين، وضحنا أنه مؤشر يهتم بقياس درجة التشتت أو انتشار مفردات البيانات حول وسطها الحسابي. وضمن هذا المفهوم، فإنه يمكن في التوزيعات الاحتمالية المشتركة قياس درجة تشتت المتغير العشوائي X_1 ودرجة تشتت المتغير العشوائي X_2 كل على حده (من خلال التوزيعات الهامشية)، كما تناولنا في الأمثلة السابقة. إضافة إلى ذلك فإنه يمكن قياس درجة التشتت بين المتغيرين وهو ما يعرف بمفهوم **التغاير**، والذي يعتمد عليه حساب مقياس إحصائي هام هو **الارتباط** بين المتغيرين كما سنرى.

تعريف (17.5): **التغاير** (بين متغيرين) (Covariance): إذا كان X_1 و X_2 متغيران عشوائيان (منفصلان أو متصلان) فإن التغاير بينهما، والذي يرمز له بالرمز $Cov(X_1,X_2)$ يعرف بالصيغة:

$$Cov(X_1, X_2) = E[(X_1-E(X_1))(X_2-E(X_2))] = E(X_1 X_2) - E(X_1)E(X_2)$$

ولاحظ أنه من النظرية (5.4) أنه إذا كان X_1 و X_2 مستقلان فإن

$$Cov(X_1, X_2) = E(X_1)E(X_2) - E(X_1)E(X_2) = 0$$

وهذا يعني عدم وجود علاقة بين X_1 و X_2 . وفي الواقع يمكن قياس الارتباط بين أي متغيرين عشوائيين منفصلين أو متصلين باستخدام معامل يعرف **بمعامل الارتباط** كما هو موضح في التعريف التالي:

تعريف (18.5): **معامل الارتباط** (بين متغيرين) (Correlation Coefficient): إذا كان X_1 و X_2 متغيران عشوائيان (منفصلان أو متصلان) لهما التغاير $Cov(X_1, X_2)$ والتباينات $Var(X_1)$ و $Var(X_2)$ على الترتيب فإن الارتباط بين X_1 و X_2 يمكن قياسه باستخدام معامل الارتباط، والذي يرمز له بالرمز $\rho_{X_1X_2}$ ، كما يلي:

$$\rho_{X_1X_2} = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)Var(X_2)}} \quad , \qquad -1 \le \rho_{X_1X_2} \le 1$$

ملاحظة: من التعريف السابق، يتضح أن قيمة معامل الارتباط $\rho_{X_1X_2}$ ستكون صفرا عندما يكون التغاير بين المتغيرين هو صفر، أي لا توجد علاقة بينهما. وكلما زادت قيمة $\rho_{X_1X_2}$ باتجاه الواحد الصحيح بالموجب كلما دل ذلك على وجود علاقة طرديه قوية بين المتغيرين، وكلما زادت قيمته باتجاه الواحد الصحيح بالسالب كلما دل ذلك على وجود علاقة عكسية قوية بين المتغيرين. ومعنى قوة العلاقة بين المتغيرين أنه كلما زاد التغير في قيم أحد المتغيرين زاد التغير في قيم المتغير الآخر سواء بالزيادة أو النقصان.

مثال (17.5): في المثال (12.5) إذا تم سحب الكرتين بالإرجاع؛

1) أوجد التوزيع الاحتمالي المشترك للمتغيرين X_1 و X_2 وكذلك الدوال الهامشية لهما.

2) أحسب معامل الارتباط بين X_1 و X_2 مع التعليق.

الحل:

1) عندما يكون السحب بالإرجاع فإن قيم المتغيرين لن تتغير ولكن ستتغير الاحتمالات المشتركة بينهما كما هو موضح في جدول (11.5).

جدول(11.5): التوزيع الاحتمالي المشترك لألوان الكرات المسحوبة بالإرجاع (مثال (12.5)).

X_1 \ X_2	0	1	المجموع
0	4/25	6/25	2/5
1	6/25	9/25	3/5
المجموع	2/5	3/5	1

ونلاحظ أن التوزيعات الهامشية للمتغيرين X_1 و X_2 هي نفسها التي في جدول (8.5).

2) لدينا

$$\mathrm{E}(X_1) = \frac{3}{5}\ ,\ \mathrm{E}(X_2) = \frac{3}{5}\ ,\ \mathrm{E}(X_1X_2) = \frac{9}{25}$$

وبالتالي فإن التغاير بين X_1 و X_2 هو

$$Cov(X_1,X_2) = \mathrm{E}(X_1X_2) - \mathrm{E}(X_1)\mathrm{E}(X_2) = \frac{9}{25} - \frac{3}{5} \times \frac{3}{5} = 0$$

وبالتالي لا داعي في هذا المثال لحساب تباينات المتغيرين لأن الارتباط بين المتغير سيساوي بدوره الصفر؛

$$\rho_{X_1X_2} = \frac{Cov(X_1,X_2)}{\sqrt{Var(X_1)Var(X_2)}} = 0$$

وهذا يعني أن X_1 و X_2 (عند تغير طريقة سحب الكرات في التجربة) قد أصبحا متغيرين مستقلين لا علاقة بينهما.

2.5.5 التوزيعات الاحتمالية المشتركة المتصلة (الثنائية)

(Joint Continuous Probability Distributions)

وهي التوزيعات التي تكون فيها المتغيرات العشوائية متغيرات متصلة لها دالة كثافة احتمالية مشتركة معرفة على فترة معينة.

تعريف (19.5): **دالة الكثافة الاحتمالية المشتركة لمتغيران متصلان:** إذا كان X_1 و X_2 متغيران عشوائيان متصلان فإن دالة الكثافة الاحتمالية المشتركة لها تعرف بالصورة:

$$f_{X_1X_2}(x_1,x_2) = f(x_1,x_2)\ , -\infty < x_1 < \infty\ \ , -\infty < x_2 < \infty$$

وتحقق الشروط

1) $f(x_1,x_2)\geq 0 \quad ,\forall x_1,x_2$.

2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2) = 1$.

تعريف (20.5): **دالة التوزيع التراكمي المشتركة لمتغيران متصلان:** إذا كان X_1 و X_2 متغيران عشوائيان متصلان لهما دالة الكثافة المشتركة $f(x_1,x_2)$ فإن دالة التوزيع التراكمية لهما تعرف بالصورة

$$F(x_1,x_2)=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f(t_1,t_2)dt_1dt_2 \quad ,\forall -\infty < x_1 < \infty \quad ,-\infty < x_2 < \infty$$

تعريف (21.5): **دالة الكثافة الاحتمالية الهامشية:** إذا كان X_1 و X_2 متغيران عشوائيان متصلان لهما دالة الكثافة المشتركة $f(x_1,x_2)$ فإن دوال الكثافة الهامشية لهما على الترتيب هي

$$f_2(x_2)=\int_{-\infty}^{\infty} f(x_1,x_2)dx_1 \quad \text{و} \quad f_1(x_1)=\int_{-\infty}^{\infty} f(x_1,x_2)dx_2$$

ملاحظات: إذا كان X_1 و X_2 متغيران عشوائيان متصلان فإنه يمكن كتابة النقاط التالية:

1. تكون دالة الاحتمال الشرطي للمتغير X_2 علما بأن $X_1 = x_1$ معرفة بالصورة
$$f(x_2\backslash x_1)=\frac{f(x_1,x_2)}{f(x_1)} \quad , \quad f(x_1)>0$$

2. إذا كان $f(x_1,x_2)=f_1(x_1).f_2(x_2)$ فإن المتغيران X_1 و X_2 يكونا مستقلين، ويتبع ذلك أن $f(x_2\backslash x_1)=f_2(x_2)$.

3. التوقع المشترك للمتغيرين X_1 و X_2 يعرف بالصورة

$$E(X_1X_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2f(x_1,x_2)dx_1dx_2$$

مثال (18.5): إذا كانت $f(x_1,x_2)$ هي دالة الكثافة الاحتمالية المشتركة للمتغيرين العشوائيين المتصلين X_1 و X_2 ، والمعرفة بالصورة؛

$$f(x_1,x_2)=\begin{cases} cx_1x_2^2\ ; & 0\leq x_1\leq 1\ ,\ 0\leq x_2\leq 1 \\ 0\ ; & \text{خلاف ذلك} \end{cases}$$

فالمطلوب:

1. إيجاد قيمة الثابت c.
2. هل المتغيرين X_1 و X_2 مستقلين؟.

الحل:

1. من شروط دالة كثافة الاحتمال المشتركة أن

$$\int_0^1 \int_0^1 c x_1 x_2^2 \, dx_1 dx_2 = 1$$

$$c \int_0^1 x_2^2 \int_0^1 x_1 \, dx_1 dx_2 = 1$$

$$c \int_0^1 x_2^2 \left[\frac{x_1^2}{2}\right]_0^1 dx_2 = 1$$

$$\frac{c}{2} \int_0^1 x_2^2 \, dx_2 = 1$$

$$\frac{c}{6} [x^3]_0^1 = 1$$

$$\frac{c}{6} = 1 \quad \rightarrow \quad c = 6$$

2. نقوم أولا بإيجاد الدوال الهامشية للمتغيرين X_1 و X_2

$$f_1(x_1) = 6 \int_0^1 x_1 x_2^2 \, dx_2 = \frac{6}{3} x_1 [x_2^3]_0^1 = 2x_1 \quad , \quad 0 \le x_1 \le 1$$

وبالمثل

$$f_2(x_2) = 6 \int_0^1 x_1 x_2^2 \, dx_1 = \frac{6}{2} x_2^2 [x_1^2]_0^1 = 3x_2^2 \quad , \quad 0 \le x_2 \le 1$$

وبضرب الدوال الهامشية للمتغيرين X_1 و X_2 بعضها ببعض نحصل على

$$f_1(x_1).f_2(x_2) = 2x_1 3x_2^2 = 6x_1 x_2^2 = f(x_1, x_2)$$

وهذا يعني أن المتغيران X_1 و X_2 مستقلان.

مثال (19.5): إذا كان Y_1 و Y_2 متغيران عشوائيان متصلان لهما دالة الكثافة المشتركة

$$f(y_1, y_2) = \begin{cases} 3y_1 \, ; & 0 \le y_2 \le y_1 \le 1 \\ 0 \, ; & \text{خلاف ذلك} \end{cases}$$

فأوجد التغاير بين Y_1 و Y_2 .

الحل:

نقوم بحساب $E(Y_1)$ ، $E(Y_1)$ ، و $E(Y_1Y_2)$ كالتالي

$$E(Y_1) = \int_0^1 \int_0^{y_2} y_1(3y_1)dy_2dy_1 = \int_0^1 3y_1^3 dy_1 = \frac{3}{4}[y_1^4]_0^1 = \frac{3}{4}$$

و

$$E(Y_2) = \int_0^1 \int_0^{y_1} y_2(3y_1)dy_2dy_1 = \int_0^1 \frac{3}{2}y_1[y_2^2]_0^{y_1}dy_1 = \int_0^1 \frac{3}{2}y_1^3 dy_1 = \frac{3}{8}[y_1^4]_0^1 = \frac{3}{8}$$

و أيضا

$$E(Y_1Y_2) = \int_0^1 \int_0^{y_1} y_1y_2(3y_1)dy_2dy_1$$
$$= \int_0^1 \frac{3}{2}y_1^2[y_2^2]_0^{y_1}dy_1 = \int_0^1 \frac{3}{2}y_1^4 dy_1 = \frac{3}{10}[y_1^5]_0^1 = \frac{3}{10}$$

وبالتالي

$$Cov(Y_1, Y_2) = E(Y_1)E(Y_2) - E(Y_1)E(Y_2) = \frac{3}{10} - \frac{3}{4} \times \frac{3}{8} = 0.02$$

وهذا يدل على وجود علاقة طردية ضعيفة بين المتغيرين.

ومن القوانين المساعدة والهامة في نظرية الاحتمال، والتي ترتكز عليها بعض المفاهيم في نظريات الاستدلال الإحصائي هو ما يعرف بقانون الأعداد الكبيرة.

تعريف (22.5): **قانون الأعداد الكبيرة** (Law of large numbers): إذا كان لدينا عدد من المتغيرات العشوائية المستقلة $X_1, X_2, ..., X_n$ والتي لها نفس التوزيع الاحتمالي ونفس التوقع $E(X)$. عندها ينص قانون الأعداد الكبيرة على أن

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n}\right| > \varepsilon\right) \to 0$$

عندما $n \to \infty$ ، لأي قيمة $\varepsilon > 0$. بمعنى أن التوقع أو الوسط الحسابي لعدد من المتغيرات العشوائية (المعرّفة على توزيع احتمالي ما) يقترب من القيمة المتوقعة لهذا التوزيع عندما يكون عدد هذه المتغيرات كبير جدا.

6.5 عزوم المتغير العشوائي (Moments of Random Variable)

ناقشنا في ما سبق، في الفصل الثاني، مفهوم العزوم وطرق حسابها حول نقطة الأصل وحول الوسط الحسابي وذلك باستخدام البيانات المفردة والمبوبة ومن ثم استخدام تلك العزوم في حساب مقاييس الالتواء والتفرطح.

وبالنسبة للمتغيرات العشوائية فإن المقاييس الاستكشافية مثل الوسط الحسابي والتباين رغم أهميتها الكبيرة، لا تزودنا عادة بوصف مميز أو وحيد للتوزيع الاحتمالي للمتغير العشوائي، فقد تكون قيمة الوسط الحسابي هي نفسها لعدة توزيعات احتمالية مختلفة، وهذا بالطبع لا يؤدي إلى استنتاج أن هذه التوزيعات هي متماثلة. لذلك فإننا نلجأ للعزوم (وللدوال المولدة للعزوم كما سنرى في الجزء القادم)، لوصف التوزيعات الاحتمالية.

وفي هذا البند من الفصل الخامس، سيتم عرض طرق حساب العزوم للمتغيرات العشوائية المنفصلة والمتصلة، وذلك من خلال استخدام الدوال الاحتمالية لتلك المتغيرات.

تعريف (23.5): **عزوم المتغير العشوائي:** إذا كان X متغير عشوائي، وكان r هو عدد صحيح موجب فإن:

1. العزم الرائي حول الصفر للمتغير X يعرّف بالصورة

$$\acute{\mu}_r = E(X^r) = \begin{cases} \sum_x x^r P(x) & : \text{إذا كان } X \text{ منفصل} \\ \int_{-\infty}^{\infty} x^r f(x) dx & : \text{إذا كان } X \text{ متصل} \end{cases}$$

2. العزم الرائي حول الوسط الحسابي للمتغير X يعرّف بالصورة

$$\mu_r = E(X - E(X))^r = \begin{cases} \sum_x (x - E(X))^r P(x) & : \text{إذا كان } X \text{ منفصل} \\ \int_{-\infty}^{\infty} (x - E(X))^r f(x) dx & : \text{إذا كان } X \text{ متصل} \end{cases}$$

مثال (20.5): إذا كان X متغير عشوائي منفصل له دالة الكتلة الاحتمالية التالية $P(x) = 1/4$ حيث $x = 2, 4, 8, 16$ فأوجد:

1. العزوم الأربعة الأولى حول الصفر
2. العزوم الأربعة الأولى حول الوسط الحسابي
3. معامل الالتواء والتفرطح لتوزيع X

الحل:

1. نبدأ بتكوين جدول التوزيع الاحتمالي للمتغير العشوائي X وذلك بالتعويض بقيم X في الدالة الاحتمالية $P(x)$ فنحصل على جدول (12.5). وبالتالي تكون قيم العزوم الأربعة الأولى حول الصفر هي:

جدول (12.5): التوزيع الاحتمالي للمتغير X (مثال(12.5)).

x	2	4	8	16
$P(x)$	1/4	1/4	1/4	1/4

$$\acute{\mu}_1 = E(X) = \sum_x x\, P(x) = 2 \times \frac{1}{4} + 4 \times \frac{1}{4} + 8 \times \frac{1}{4} + 16 \times \frac{1}{4} = \frac{30}{4} = 7.5$$

$$\acute{\mu}_2 = E(X^2) = \sum_x x^2 P(x) = 2^2 \times \frac{1}{4} + 4^2 \times \frac{1}{4} + 8^2 \times \frac{1}{4} + 16^2 \times \frac{1}{4} = \frac{340}{4} = 85$$

$$\acute{\mu}_3 = E(X^3) = \sum_x x^3 P(x) = 2^3 \times \frac{1}{4} + 4^3 \times \frac{1}{4} + 8^3 \times \frac{1}{4} + 16^3 \times \frac{1}{4} = \frac{4680}{4} = 1170$$

$$\acute{\mu}_4 = E(X^4) = \sum_x x^4 P(x) = 2^4 \times \frac{1}{4} + 4^4 \times \frac{1}{4} + 8^4 \times \frac{1}{4} + 16^4 \times \frac{1}{4} = \frac{69904}{4} = 17476$$

2. باستخدام علاقة العزوم حول الوسط الحسابي بالعزوم حول الصفر، (راجع الفصل الثاني، البند 6.2)، نحصل على العزوم الأربعة الأولى حول الوسط الحسابي لتوزيع X بالصورة:

$$\mu_1 = E\big(X - E(X)\big) = E(X) - E(X) = 0$$

$$\mu_2 = E\big(X - E(X)\big)^2 = \acute{\mu}_2 - \acute{\mu}_1^2 = Var(X) = 85 - (7.5)^2 = 28.75$$

$$\mu_3 = \acute{\mu}_3 - 3\acute{\mu}_1\acute{\mu}_2 + 2\acute{\mu}_1^3 = 1170 - 3 \times 85 \times 7.5 + 2 \times (7.5)^3 = 101.25$$

$$\mu_4 = \acute{\mu}_4 - 4\acute{\mu}_1\acute{\mu}_3 + 6\acute{\mu}_2\acute{\mu}_1^2 - 3\acute{\mu}_1^4 = 17476 - 4 \times 1170 \times 7.5 + 6 \times 85 \times (7.5)^2 - 3 \times (7.5)^4 = 1571.31$$

3. لحساب معاملي الالتواء والتفرطح لتوزيع X نستخدم معامل الالتواء العزمي

$$SK_\mu = \frac{\mu_3^2}{(SD(X))^3} = \frac{\mu_3^2}{\mu_2^3} = \frac{(101.25)^2}{(28.75)^3} = 0.43$$

وهذا يعني أن توزيع X له التواء بسيط موجب (ناحية اليسار).

ومعامل التفرطح العزمي

$$Kur_\mu = \frac{\mu_4}{(SD(X))^4} = \frac{\mu_4}{\mu_2^2} = \frac{1571.31}{(28.75)^2} = 1.90$$

مما يدل على أن توزيع المتغير X هو مفرطح.

مثال (21.5): إذا كان X متغير عشوائي متصل له دالة الكثافة الاحتمالية التالية:

$$f(x) = \begin{cases} 2x\,; 0 < x < 1 \\ 0\,; \quad \text{خلاف ذلك} \end{cases}$$

فأوجد العزوم الأربعة الأولى حول الصفر.

الحل:

$$\acute{\mu}_1 = E(X) = \int_0^1 x(2x)\,dx = 2\int_0^1 x^2\,dx = \tfrac{2}{3}[x^3]_0^1 = \frac{2}{3}$$

$$\acute{\mu}_2 = E(X^2) = \int_0^1 x^2(2x)\,dx = 2\int_0^1 x^3\,dx = \tfrac{2}{4}[x^4]_0^1 = \frac{1}{2}$$

$$\acute{\mu}_3 = E(X^3) = \int_0^1 x^3(2x)\,dx = 2\int_0^1 x^4\,dx = \tfrac{2}{5}[x^5]_0^1 = \frac{2}{5}$$

$$\acute{\mu}_4 = E(X^4) = \int_0^1 x^4(2x)\,dx = 2\int_0^1 x^5\,dx = \tfrac{2}{6}[x^6]_0^1 = \frac{1}{3}$$

7.5 الدالة المولدة للعزوم والدالة المميزة

(Moment Generating Function and Characteristic Function)

رأينا في الجزء السابق (6.5) كيفية التعامل مع العزوم للمتغير العشوائي في الحالتين المنفصلة والمتصلة. وتم توضيح كيفية الحصول على العزم الرائي بالتعويض عن القيمة المرغوبة لـ r . وفي هذا الجزء سنستعرض ما يعرف بالدالة المولدة للعزوم والتي "تلخّص" إن صح التعبير كل عزوم المتغير العشوائي في صيغة رياضية واحدة تعبر عن التوزيع الاحتمالي للمتغير.

تعريف (24.5): **الدالة المولدة للعزوم** (Moment Generating Function, MGF): الدالة المولدة للعزوم، والتي يرمز لها بالرمز $\mu_x(t)$ ، للمتغير العشوائي X (المنفصل أو المتصل) تعرف بالصورة:

$$\mu_x(t) = E(e^{tX}) = 1 + \tfrac{t}{1!}E(X) + \tfrac{t^2}{2!}E(X^2) + \tfrac{t^3}{3!}E(X^3) + \cdots$$

$$= 1 + \tfrac{t}{1!}\acute{\mu}_1 + \tfrac{t^2}{2!}\acute{\mu}_2 + \tfrac{t^3}{3!}\acute{\mu}_3 + \cdots$$

ملاحظة: من خلال تعريف التوقع نستطيع كتابة الدالة المولدة للعزوم بالصورة:

$$\mu_x(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tX}P(x) & : \text{إذا كان } X \text{ منفصل} \\ \int_{-\infty}^{t} e^{tX}f(x)dx & : \text{إذا كان } X \text{ متصل} \end{cases}$$

نظرية (8.5): إذا وجدت الدالة المولدة للعزوم $\mu_x(t)$ لمتغير عشوائي X ، فيكون لأي قيمة صحيحة موجبة r يكون:

$$\left.\frac{\partial^r \mu_x(t)}{\partial t^r}\right|_{t=0} = \mu_x^r(0) = \acute{\mu}_r = E(X^r)$$

النظرية السابقة تعني أنه إذا ما تم إيجاد التفاضل رقم r للدالة المولدة للعزوم $\mu_x(t)$ بالنسبة لـ t، ثم تم وضع $t=0$، فإننا نحصل على (نولّد) العزم $\acute{\mu}_r$ ، ولهذا سميت $\mu_x(t)$ بالدالة المولدة للعزوم. فعلى سبيل المثال يمكننا إيجاد العزمين الأول والثاني حول الصفر باستخدام الدالة المولدة للعزوم كما يلي:

$$\acute{\mu}_1 = \frac{\partial \mu_x(t)}{\partial t} = \frac{\partial E(e^{tX})}{\partial t} = E\frac{\partial e^{tX}}{\partial t} = E(Xe^{tX})|_{t=0} = E(X)$$

وكذلك فإن

$$\acute{\mu}_2 = \frac{\partial^2 \mu_x(t)}{\partial t^2} = E\left(\frac{\partial (Xe^{tX})}{\partial t}\right)\Bigg|_{t=0} = E(X^2)$$

ويمكن ملاحظة أن الدالة المولدة للعزوم حول الوسط الحسابي ستكون بالصيغة:

$$\mu_{X-E(X)}(t) = E\left(e^{t(X-E(X))}\right) = 1 + \tfrac{t}{1!}E(X-E(X)) + \tfrac{t^2}{2!}E(X-E(X))^2 + \cdots$$
$$= \mu_x(t).\, e^{-tE(X)}$$

نظرية (9.5): إذا كان X وY متغيران عشوائيان لهما نفس الدالة المولدة للعزوم، وذلك لكل قيم t في الفترة حول النقطة $t = 0$ ، فإن التوزيع الاحتمالي لكل من X و Y يكون متطابقا، بمعنى أن الدالة المولدة للعزوم هي دالة وحيدة.

مثال (22.5): في المثال (20.5) أوجد الدالة المولدة للعزوم للمتغير X ثم أوجد العزمين الأول والثاني حول الصفر باستخدامها.

الحل:

$$\mu_x(t) = E(e^{tX}) = \sum_x e^{tX}P(x) \;= e^{2t}\times\frac{1}{4} + e^{4t}\times\frac{1}{4} + e^{8t}\times\frac{1}{4} + e^{16t}\times\frac{1}{4}$$
$$= \frac{1}{4}(e^{2t} + e^{4t} + e^{8t} + e^{16t})$$

ولإيجاد العزم الأول حول الصفر نوجد التفاضل الأول للدالة $\mu_x(t)$

$$\acute{\mu}_1 = \frac{\partial \mu_x(t)}{\partial t}\Bigg|_{t=0} = \frac{1}{4}(2e^{2t} + 4e^{4t} + 8e^{8t} + 16e^{16t})\Bigg|_{t=0} = \frac{1}{4}(2+4+8+16)$$
$$= 7.5$$

$$\acute{\mu}_2 = \frac{\partial^2 \mu_x(t)}{\partial t^2}\Bigg|_{t=0} = \frac{1}{4}(2\times 2e^{2t} + 4\times 4e^{4t} + 8\times 8e^{8t} + 16\times 16e^{16t})\Bigg|_{t=0}$$
$$= \frac{1}{4}(4+16+64+256) = 85$$

وسنتطرق لمزيد من الدوال المولدة للعزوم وتطبيقاتها بعد دراسة بعض أشهر أنواع التوزيعات الاحتمالية المنفصلة والمتصلة في الفصل القادم.

من الناحية الرياضية، فإن الدالة المولدة للعزوم قد لا يمكن حسابها في بعض الحالات أو لبعض التوزيعات الاحتمالية. لذلك يمكن اللجوء لدالة أخرى هي **الدالة المميزة**. وقد سميت بهذا الاسم لأنها "تُميز" توزيع المتغير العشوائي أو التوزيع الاحتمالي عن باقي التوزيعات. وهذه الدالة يمكن دائما الحصول عليها لأن دوال التوزيع الاحتمالي تكون قابلة للتكامل رياضيا. إضافة إلى ذلك، فإن الدالة المميزة تساوي في الواقع معكوس **تحويل فوريه**[1] (Fourier Transformation) لدالة التوزيع الاحتمالي.

تعريف (25.5): **الدالة المميزة** (Characteristic Function): الدالة المميزة، والتي يرمز لها بالرمز $\varphi_x(t)$، للمتغير العشوائي X (المنفصل أو المتصل) ولأي قيمة حقيقية t ، والجذر التخيُلي $i=\sqrt{-1}$ ، تعرف بالصورة:

$$\varphi_x(t)=E\left(e^{itX}\right)=\int_{-\infty}^{\infty} e^{itx}\, f(x)dx$$

ونلاحظ من التعريف السابق أن $\varphi_x(t)=\mu_x(it)$ ، أي أن الدالة المميزة تساوي الدالة المولدة للعزوم بمعلمة it . وسنكتفي بالمثال السابق على الدالة المولدة للعزوم دون إدراج مثال على الدالة المميزة، (إضافة لما سنتناوله من تطبيقات الدالة المولدة للعزوم في الفصل القادم)، حيث أن تطبيقات الدالة المميزة تتجاوز المستوى الحالي في دراسة التوزيعات الاحتمالية.

[1] إذا كان X متغير عشوائي له الدالة الاحتمالية $f(x)$ فإن التكامل $\int_{-\infty}^{\infty} e^{-itx}\, f(x)dx$ يُعرف **بتحويل فوريه**، حيث $i=\sqrt{-1}$.
وهذا التحويل كثيرا ما يستخدم في التعامل مع التوزيعات الاحتمالية المعقدة، ويمكن استخدامه مع أي دالة قابلة للتكامل.

5.8 تمارين الفصل الخامس

تمرين (1.5): أحد أقسام معرض الكتاب به رف يحوي 15 كتاب منها 4 كتب أطفال ، فإذا تم اختيار 4 كتب من هذا الرف عشوائيا، وكان X متغير عشوائي يمثل عدد الكتب المُختارة فأوجد:

1. التوزيع الاحتمالي لعدد كتب الأطفال المختارة.
2. احتمال عدم وجود أي كتاب للأطفال ضمن الكتب المختارة.
3. احتمال وجود كتاب واحد للأطفال على الأكثر ضمن الكتب المختارة.
4. احتمال وجود ثلاثة كتب أطفال على الأقل ضمن الكتب المختارة.
5. احتمال وجود من كتابين إلى أربعة كتب أطفال ضمن الكتب المختارة.
6. F(2) .
7. أرسم المدرج الاحتمالي لقيم المتغير العشوائي X.

تمرين (2.5): إذا كان Y متغير عشوائي متصل له الدالة التالية:

$$f(y) = \begin{cases} \frac{1}{9}\,Y^2\,; 0 \le y \le 3 \\ 0\ ;\ \text{خلاف ذلك} \end{cases}$$

1. أثبت أن $f(y)$ هي دالة كثافة احتمالية.
2. أوجد $P(Y > 2)$.
3. أوجد $P(0 \le Y \le 2)$.
4. أوجد $P(2 \le Y \le 5)$
5. أوجد $F(1)$ و $F(3)$.

تمرين (3.5): إذا كان X متغير عشوائي متصل له دالة الكثافة الاحتمالية التالية:

$$f(x) = \begin{cases} aX^2 + \frac{1}{9}X\,;\ \ 0 \le x \le 3 \\ 0\ ;\ \text{خلاف ذلك} \end{cases}$$

فأوجد قيمة الثابت a.

تمرين (4.5): للتوزيع الاحتمالي التالي أوجد:

1. قيمة k 2. $E(X)$ 3. $E(Z)$ حيث $Z = 2X - \frac{1}{2}$ 4. $Var(X)$ 5. $Var(Y)$ حيث $Y = \frac{X}{3} + 5$

x	0	1	2	3
$P(x)$	1/4	1/4	1/3	k

تمرين (5.5): باستخدام دالة الكثافة الاحتمالية في تمرين (3.5) أوجد:

1. $E(X)$ 2. $E(3X)$ 3. $Var(X)$ 4. $Var(2X - 3)$

تمرين (6.5): أحد الرفوف في متجر خضروات تبقى به 4 حبات تفاح و3 حبات برتقال. تم اختيار حبتين عشوائيا من هذا الرف بدون إرجاع، ثم تم تعريف المتغيران التاليان:

$$X_1 = \begin{cases} 0 & \text{إذا كانت الحبة الأولى المختارة هي برتقالة} \\ 1 & \text{إذا كانت الحبة الأولى المختارة هي تفاحة} \end{cases}$$

$$X_2 = \begin{cases} 0 & \text{إذا كانت الحبة الثانية المختارة هي برتقالة} \\ 1 & \text{إذا كانت الحبة الثانية المختارة هي تفاحة} \end{cases}$$

1. أوجد التوزيع الاحتمالي المشترك للمتغيرين.
2. أوجد دوال التوزيع الهامشية لكل منهما.
3. هل المتغيران مستقلان؟.
4. أوجد $E(X_1)$ ، $E(X_2)$ و $E(X_1X_2)$.

تمرين (7.5): باستخدام التوزيع الاحتمالي المشترك التالي:

X_1 \ X_2	0	1	المجموع
0	3/20	3/20	6/20
1	12/20	2/20	14/20
المجموع	15/20	5/20	1

1. أوجد $E(X_1)$ ، $E(X_2)$ و $E(X_1X_2)$.
2. أوجد معامل الارتباط بين المتغيرين.

تمرين (8.5): إذا كانت $f(x_1, x_2)$ هي دالة الكثافة الاحتمالية المشتركة للمتغيرين العشوائيين المتصلين X_1 و X_2 ، والمعطاة بالصيغة:

$$f(x_1, x_2) = \begin{cases} 2kx_1x_2^2 \,; & 0 \le x_1 \le 1 \,,\ 0 \le x_2 \le 1 \\ 0 \,; & \text{خلاف ذلك} \end{cases}$$

أوجد قيمة الثابت k .

تمرين (9.5): باستخدام التوزيع الاحتمالي المنفصل التالي:

x	-3	-1	0	1	3
$P(x)$	1/5	1/5	1/5	1/5	1/5

أوجد

1. العزوم الأربعة الأولى حول الصفر.
2. العزوم الأربعة الأولى حول الوسط الحسابي.
3. معاملي الالتواء والتفرطح لتوزيع المتغير العشوائي.
4. الدالة المولدة للعزوم للتوزيع الاحتمالي.
5. العزم الأول والثاني حول الصفر باستخدام الدالة المولدة للعزوم.

الفصل السادس

التوزيعات الاحتمالية المنفصلة والمتصلة
(Discrete and Continuous Probability Distributions)

1.6 مقدمة (Introduction)

ناقشنا في الفصل السابق مفهوم المتغيرات العشوائية بنوعيها المنفصل والمتصل وتوزيعاتها الاحتمالية، وكذلك الخواص المتعلقة بها وكيفية حساب بعض المقاييس الوصفية لها مثل الوسط الحسابي والتباين. وقد لاحظنا في بعض التجارب العشوائية أن كثيرا من المتغيرات قد تسلك نفس السلوك، بمعنى أن طبيعة التجربة قد تفرض أو تحدد طبيعة القيم التي سيأخذها المتغير العشوائي وطريقة حساب الاحتمالات المناظرة لهذه القيم.

وهكذا، فإن هذه المتغيرات العشوائية الملازمة أو المُعرفة على هذه التجارب يمكن أن يتم حساب احتمالات حدوثها باستخدام نفس التوزيع الاحتمالي، بمعنى أنه يمكن إيجاد توزيع احتمالي واحد يمثل طبيعة التغير الواقع، ويمكن في نفس الوقت حساب الاحتمالات المناظرة لقيم المتغير بشكل مباشر. في الحقيقة، توجد عدة توزيعات احتمالية منفصلة ومتصلة مُعدّة بشكل نماذج أو دوال احتمالية بحيث يتم تصنيف سلوك المتغيرات، (بحسب طبيعة التجربة العشوائية)، طبقا لتلك الدوال وبالتالي الحصول على قيم المتغير والاحتمالات المناظرة بصورة مباشرة وسريعة.

لذلك فإنه يمكن اعتبار هذا الفصل كمرجع لمساعدة القارئ في التعرف على أهم التوزيعات الاحتمالية المنفصلة والمتصلة والتي تعتبر أدوات أساسية في التعامل مع طرق ومفاهيم الاستدلال الإحصائي. وسنبدأ بالتوزيعات الاحتمالية المنفصلة في البند (2.6)، ثم تليها التوزيعات المتصلة في البند (3.6)، حيث سيتم تعريف كل توزيع احتمالي واستعراض خواصه والتعرف على التوقع والتباين له، وكذلك دالته المولدة للعزوم.

2.6 بعض التوزيعات الاحتمالية المنفصلة (Some Discrete Probability Distributions)

1.2.6 التوزيع المنتظم المنفصل (Discrete Uniform Distribution)

وهو يعد من أبسط التوزيعات الاحتمالية، ويتميز بأنه يأخذ الشكلين المنفصل والمتصل[1] تبعا لطبيعة التجربة وتعريف المتغير العشوائي.

تعريف (1.6): **التوزيع المنتظم المنفصل** (Discrete Uniform Distribution): إذا كان X متغير عشوائي منفصل يأخذ القيم $x_1, x_2, \dots, x_k$ باحتمالات متساوية، فإن التوزيع المنتظم المنفصل يعرّف بالصورة:

$$P(x;k) = \frac{1}{k} \quad , \quad x = x_1, x_2, \dots, x_k$$

ويقال أن المتغير العشوائي X يتبع التوزيع المنتظم المنفصل بمعلمة k ، ويرمز لذلك بالصورة $X \sim Uniform\,(x;k)$.

ملاحظة: سيتم ابتداء من هذا الفصل استخدام الشكل القياسي ($f(x;\theta)$ أو $P(x;\theta)$) كرمز للتوزيعات الاحتمالية المنفصلة أو المتصلة لأن هذا الشكل يوضح للقارئ معلمة التوزيع θ والتي يتم تعريفها بالصورة التالية:

---

[1] سيتم تعريف التوزيع المنتظم المتصل في البند (3.6).

تعريف (2.6): **معلمة التوزيع الاحتمالي** (Parameter of the probability distribution): القيمة التي تحدد الشكل المميز لدالة التوزيع الاحتمالي تعرّف بأنها معلمة التوزيع الاحتمالي.

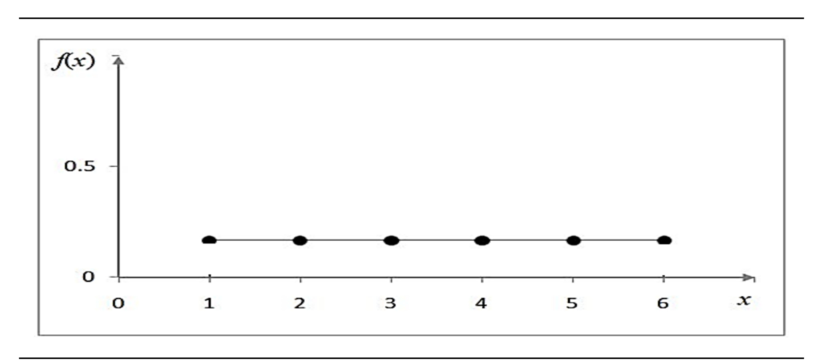
وتجدر الإشارة هنا إلى أن الفرق بين القيمة الثابتة (Constant value) والمعلمة في مفهوم الاحتمالات أن المصطلح الأول يقصد به أي قيمة لا تتغير من توزيع احتمالي لآخر أو ضمن التوزيع نفسه (مثل الثابت $\pi = 3.14$ مثلا)، أما المعلمة فهي قيمة تكون ثابتة عند تعريف المتغير (أو التوزيع الاحتمالي) بعد إجراء التجربة العشوائية، وتتغير بتغير هذا التعريف سواء في نفس التجربة أو في تجارب أخرى، كما سنلاحظ من خلال دراستنا للتوزيعات المختلفة، وينطبق هذا المفهوم على نوعي التوزيعات المنفصلة والمتصلة.

مثال (1.6): في تجربة إلقاء زهر نرد، يكون فراغ العينة $S = \{1, 2, 3, 4, 5, 6\}$ ، ونلاحظ أن كل قيمة من قيم المتغير العشوائي X الستة، (والذي يمثل الرقم الظاهر على زهر النرد)، يكون لها نفس فرصة الحدوث، أي أنه لها نفس احتمال الظهور وهذا يعني أن المتغير X في هذه الحالة هو متغير عشوائي يتبع التوزيع المنتظم المنفصل. وبشكل رياضي نستطيع التعبير عن ذلك بالصورة $X \sim Uniform\ (x; 6)$.

حيث تكون دالة الكتلة الاحتمالية للتوزيع هي

$$P(x; k) = P(x; 6) = \frac{1}{6} \quad , \quad x = 1, 2, 3, 4, 5, 6$$

ويمكن ملاحظة سبب تسمية التوزيع المنتظم بهذا الاسم من خلال الرسم البياني للتوزيع كما هو موضح في الشكل (1.6)، حيث يلاحظ أن الاحتمالات المناظرة لكل قيم المتغير X هي متساوية وكأنها تسير في خط منتظم.



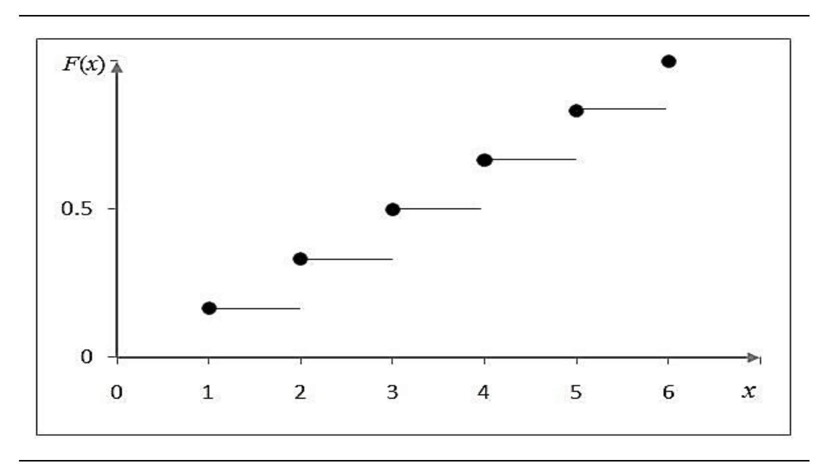
شكل (1.6): التوزيع الاحتمالي للمتغير X من المثال (1.6).

وكذلك يمكن إيجاد دالة التوزيع التراكمي للمتغير X ويرمز لها بالرمز $F(x;k)$ كما هو مبين في جدول (1.6).

جدول (1.6): التوزيع الاحتمالي التراكمي للمتغير X من المثال (1.6).

x	1	2	3	4	5	6
$F(x;6)$	1/6	2/6	3/6	4/6	5/6	6/6

وأيضا تمثيلها بيانيا كما يوضح الشكل (2.6).



شكل (2.6): المدرج الاحتمالي المتصاعد (التراكمي) للمتغير X من المثال (1.6).

مثال (2.6): أوجد التوزيع الاحتمالي للمتغير العشوائي الذي يمثل الخانة الأولى للرقم التسلسلي الموجود على سلعة منتجة من أحد المصانع. ثم أوجد احتمال الحصول على الرقم 3 من بينها.

الحل:

حيث أن الخانة الأولى (أو أي خانة أخرى) ضمن الرقم التسلسلي لأي سلعة منتجة ستحتوي على رقم من 0 إلى 9 فيكون فراغ العينة لهذه التجربة هو $S = \{0, 1, 2, 3, 4, ..., 9\}$ ، وبالتالي فإن احتمال اختيار أي رقم من هذا الفراغ سيكون 1/10 وبالتالي فإن:

$$X \sim Uniform\ (x;10)$$

$$P(x;k) = P(x;10) = \frac{1}{10} \quad , \quad x = 0,1,2,3,\dots,9$$

ويقال أيضا أن المتغير X معرّف على الفترة [9 , 0] والتي تضم قيما منفصلة. ويكون احتمال اختيار الرقم 3 (والذي يساوي احتمال اختيار أي رقم في الواقع) هو

$$P(3,10) = P(3) = 1/10$$

تعريف (3.6): **التوقع والتباين للتوزيع المنتظم المنفصل:** إذا كان X متغير عشوائي منفصل يتبع التوزيع المنتظم ويأخذ القيم $a, a+1, a+2, \dots, b$ حيث a و b هما القيمتين الصغرى والكبرى لقيم X، فإن توقع وتباين[1] المتغير X يعرّفان بالصورة:

$$E(X) = \frac{a+b}{2} \quad , \quad Var(X) = \frac{(b-a)^2}{12}$$

مثال (2.6): أوجد التوقع والتباين للمتغير العشوائي X والذي يمثل فصيلة دم الإنسان.

الحل:

[1] يمكن الوصول لهذه الصيغ بسهولة من خلال التعويض عن دالة التوزيع المنتظم في القوانين العامة لحساب التوقع والتباين للمتغير العشوائي المنفصل، (راجع الفصل الخامس).

من هذه التجربة العشوائية، والتي تمثل نتيجة الكشف عن فصيلة دم الإنسان في المختبر، يكون فراغ العينة $S =$ {A, B, AB, O} ، والذي يمكن إعادة كتابته باستخدام رموز رقمية للفصائل على النحو $S = \{1, 2, 3, 4\}$، وهكذا فإن المتغير X يكون له توزيع منتظم بالصورة:

$$P(x;k) = P(x;4) = \frac{1}{4} \quad , \quad x = 1, 2, 3, 4$$

ويكون

$$E(X) = \frac{a+b}{2} = \frac{1+4}{2} = 2.5 \quad , Var(X) = \frac{(b-a)^2}{12} = \frac{(4-1)^2}{12} = 0.75$$

مع ملاحظة أنه لا يمكن تفسير قيم التوقع والتباين بصورة عملية في هذا المثال، حيث أن $E(X) = 2.5$ مثلا لا يعني أن فصائل الدم ستتمركز ما بين الفصيلتين B و AB بالضرورة.

أما بالنسبة للدالة المولدة للعزوم للتوزيع المنتظم فسيتم التطرق إليها عند تناول التوزيع المنتظم المتصل في البند (3.6).

2.2.6 محاولات بيرنوللي وتوزيع ذي الحدين (Bernoulli Trials and Binomial Distribution)

وهو من التوزيعات الهامة والكثيرة الاستخدام لأن كثيرا من التجارب العشوائية في الحياة يمكن أن تتضمن متغيرات تسلك سلوكا يندرج تحت مفهوم توزيع ذي الحدين كما توضح لنا الأمثلة العملية التالية:

1. رمي عملة معدنية 10 مرات واعتبار المتغير العشوائي X يمثل عدد الصور (H) الظاهرة.
2. آلة في أحد المصانع تنتج 1% وحدة معيبة من إنتاجها الكلي و X يمثل عدد الوحدات المُعابة في 25 وحدة منتجة تم اختيارها من إنتاج الآلة.
3. اختبار متعدد الاختيارات (MCQ) يحتوي على 10 أسئلة، وكل سؤال له 4 إجابات، و X يمثل عدد الأسئلة التي تمت إجابتها بشكل صحيح.
4. في الولادات الـ 20 القادمة في أحد المستشفيات، X يمثل عدد المواليد الإناث.

هذه الأمثلة، تمثل سلسلة من المحاولات العشوائية المتكررة (Repeated random trials)، فالمحاولات تكررت 10 مرات عند رمي العملة في المثال الأول، وتكررت 25 مرة في إنتاج الآلة في المثال الثاني، وهكذا في باقي الأمثلة. والملاحظ أن المتغير العشوائي X في كل حالة يعبّر عن عدد المحاولات التي تحقق ظاهرة معينة. والنتيجة في كل محاولة إما أن تتوافق مع الظاهرة التي يمثلها X أو لا تتوافق.

ففي المثال الأول إما أن نحصل على صورة في المحاولة الواحدة (عند رمي العملة) أو لا نحصل عليها، وفي المثال الثاني إما أن نصادف وحدة معيبة أو لا نصادفها، وهكذا. وبالتالي يمكن اعتبار أن المحاولات في كل تجربة من التجارب السابقة يمكن تلخيصها من خلال نتيجتين متضادتين؛ إما "النجاح (Success)" في الحصول على نتيجة تتوافق مع تعريف المتغير X أو "الفشل (Failure)" في ذلك.

وبالنسبة للتجارب التي تتضمن أكثر من نتيجتين ضمن المحاولة الواحدة، كما في المثال الثالث (الاختبار متعدد الاختيارات)، فإنه يمكن اعتبار الحصول على إجابة صحيحة من ضمن الإجابات الأربع هو النجاح، وعدم الحصول على إجابة صحيحة (والذي يعني الحصول على أي إجابة من الإجابات الثلاث الخاطئة الباقية) هو الفشل.

إن التجارب العشوائية التي تتضمن محاولة لها نتيجتين ممكنتين فقط تعرف باسم تجارب أو محاولات بيرنوللي (Bernoulli Trials) نسبة إلى العالم الإيطالي الشهير. ويذهب بعض الكتاب إلى تسمية توزيع المتغير العشوائي الذي يصف تلك المحاولات بتوزيع بيرنوللي (Bernoulli Distribution)، وسوف نتناول لاحقا تكرار محاولات بيرنوللي لعدد محدد من المرات (والذي سيمثل توزيع "ذي الحدين"، على اعتبار أن الحدين هما حد النجاح وحد الفشل). أما الآن فسوف نلقي الضوء على توزيع بيرنوللي ونتعرف على خصائصه.

تعريف (4.6): **توزيع (أو محاولات) بيرنوللي:** إذا كان X متغير عشوائي معرّف على تجربة عشوائية تحقق الشروط التالية:

1. لكل محاولة في التجربة العشوائية توجد فقط نتيجتين متضادتين (نجاح أو فشل).
2. احتمالات حدوث نتائج التجربة هي ثابتة لكل محاولة.
3. المحاولات تكون مستقلة عن بعضها البعض.

فإن X عندها يتبع توزيع يعرف بتوزيع بيرنوللي ويأخذ الصيغة التالية:

$$X \sim Bernoulli\ (x; p)$$

أو

$$P(x;p) = P(X = x) = p^x(1-p)^{1-x} \quad , \quad x = 0, 1$$

حيث p (معلمة التوزيع) هو احتمال النجاح في أي محاولة من محاولات التجربة،

و $\sum_{x=0}^{1} Bernoulli(x;p) = 1$ ، حيث $0 \le p \le 1$.

مثال (3.6): في تجربة إلقاء زهر نرد يكون فراغ العينة $S = \{1, 2, 3, 4, 5, 6\}$ ، فإذا ما تم تعريف المتغير العشوائي X بأنه يمثل ظهور عدد فردي، فأوجد:

1. التوزيع الاحتمالي للمتغير X . 2. $P(X =$ عدد زوجي$)$.

الحل:

1. من تعريف المتغير العشوائي X على التجربة نرى بأنه إما أن يمثل النتيجة (ظهور عدد فردي) ويمكن تسميته بالنجاح، أو لا يأخذها (وذلك عند ظهور عدد زوجي) وتسمى النتيجة عندها بالفشل. وبالتالي تكون قيم المتغير X:

عدم ظهور عدد فردي $\rightarrow$ $X = 0$: حالة الفشل

ظهور عدد فردي $\rightarrow$ $X = 1$: حالة النجاح

وتكون معلمة التوزيع والتي تمثل احتمال الحصول على عدد فردي هي $p = 3/6 = 1/2$ ، بالتالي يكون التوزيع الاحتمالي للمتغير X:

$$X \sim Bernoulli\ (x; p) = X \sim Bernoulli\ (x;1/2)$$

أو

$$P\left(x;\tfrac{1}{2}\right) = P(X = x) = (\tfrac{1}{2})^x(\tfrac{1}{2})^{1-x} \quad , \quad x = 0, 1$$

ولاحظ أن

$$\sum_{x=0}^{1} Bernoulli\left(x;\tfrac{1}{2}\right) = Bernoulli\left(0,\tfrac{1}{2}\right) + Bernoulli\left(1,\tfrac{1}{2}\right) = \tfrac{1}{2} + \tfrac{1}{2} = 1$$

2. بالتعويض عن $x = 0$ في التوزيع الاحتمالي نحصل على

$$P(X = \text{عدد زوجي}) = P(X = 0) = (\tfrac{1}{2})^0(\tfrac{1}{2})^{1-0} = \tfrac{1}{2}$$

تعريف (5.6): **توقع وتباين توزيع بيرنوللي:** إذا كان X متغير عشوائي يتبع توزيع بيرنوللي بالمعلمة p، فإن توقعه وتباينه يعرفان بالصورة

$$E(X) = p \quad , \quad Var(X) = p(1 - p)$$

ملاحظة: يمكن الوصول لقيم التوقع والتباين في التعريف السابق كالتالي:

$$E(X) = \sum_x xP(x) = \sum_{x=0}^{1} xp^x(1-p)^{1-x} = (0)p^0(1-p)^{1-0} + (1)p^1(1-p)^{1-1} = p$$

$$E(X^2) = \sum_x x^2P(x) = \sum_{x=0}^{1} x^2p^x(1-p)^{1-x}$$
$$= (0)^2p^0(1-p)^{1-0} + (1)^2p^1(1-p)^{1-1} = p$$

وبالتالي

$$Var(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p)$$

تعريف (6.6): **الدالة المولدة للعزوم لتوزيع بيرنوللي:** إذا كان $X \sim Bernoulli\ (x; p)$ فإن الدالة المولدة للعزوم له تعرف بالصورة:

$$\mu_x(t) = E(e^{tX}) = (1-p) + pe^t$$

ملاحظة: يمكن إثبات الصيغة السابقة (تعريف (6.6)) بالصورة:

$$\mu_x(t) = E(e^{tX}) = \sum_{x=0}^{1} e^{tx}P(x) = e^{(0)t}.p^0(1-p)^{1-0} + e^{(1)t}.p^1(1-p)^{1-1}$$
$$= (1-p) + pe^t$$

مثال (4.6): إذا كان $X \sim Bernoulli\ (x; p)$ فأوجد توقعه وتباينه باستخدام الدالة المولدة للعزوم.

الحل:

لدينا من تعريف (6.6)

$$\mu_x(t) = E(e^{tX}) = (1-p) + pe^t$$

$$\acute{\mu}_1 = \frac{\partial \mu_x(t)}{\partial t} = pe^t\Big|_{t=0} = p = E(X)$$

$$\acute{\mu}_2 = \frac{\partial^2 \mu_x(t)}{\partial t^2} = pe^t\Big|_{t=0} = p = E(X^2)$$

$$\mu_2 = \acute{\mu}_2 - \acute{\mu}_1^2 = p - p^2 = p(1-p) = Var(X)$$

نأتي الآن لتعريف توزيع ذي الحدين والذي وضحنا سابقا بأنه ينتج عن تكرار محاولة بيرنوللي لعدد محدد من المرات.

تعريف (7.6): **توزيع ذي الحدين** (Binomial distribution): إذا ما تم إجراء عدد n محاولة من محاولات بيرنوللي (بالشروط الموضحة في التعريف (4.6))، وتم تعريف المتغير العشوائي X بأنه يمثل عدد مرات النجاح في الـ n محاولة، فإن توزيع المتغير X يعرّف بتوزيع ذي الحدين ويعطى بالصيغة:

$$X \sim Binomial\ (x, n, p)$$

$$P(x; n, p) = P(X = x) = C_x^n p^x (1-p)^{n-x} \quad , \quad x = 0, 1, 2, \dots, n$$

حيث n, p هي معالم توزيع ذي الحدين؛ $0 \le p \le 1$ ، و $n = 1, 2, 3, \dots$ ، و p يمثل احتمال النجاح في أي محاولة.

ويمكن "تحليل" أي متغير عشوائي X والمعرّف على أي تجربة عشوائية لمعرفة إمكانية أن يكون له توزيع يتبع ذي الحدين كما هو الحال في المثال التوضيحي التالي:

لنفرض أن هنالك عدد من الطلبة في إحدى الجامعات الأوروبية وتم اختيار 40 طالب مستجد منهم بصورة عشوائية، وتم إجراء اختبار (تحليل دم) خاص بفيروس الإيدز (HIV) بحيث توضع النتيجة "إيجابي (+ve) " للطالب المصاب، والنتيجة "سلبي (-ve)" للطالب الغير مصاب. فكانت النتيجة أن 5 طلبة من الأربعين طالب هم مصابون بالفيروس. فإذا ما تم تعريف المتغير العشوائي X بأنه يمثل عدد الطلبة المصابين بالفيروس في العينة المسحوبة (الأربعون طالب)، فهل يمكن القول أن X يتبع توزيع ذي الحدين؟.

للإجابة على هذا التساؤل يجب التأكد من أن المتغير X يستوفي الشروط المحددة في التعريف (7.6) . من التجربة العشوائية (تحليل الدم للطلبة) نستخلص التالي:

1. التجربة تتضمن $n = 40$ محاولة محددة (اختيار 40 طالب من مجتمع الطلبة).
2. لكل محاولة نتيجتان متضادتان، إما إيجابي أو سلبي.
3. احتمال إصابة أي طالب في العينة هو ثابت ويساوي $p = 5/40 = 0.125$.
4. اختبار كل طالب (تحليل الدم) تم بصورة مستقلة عن الآخر وبالتالي فإن المحاولات (محاولات بيرنوللي) مستقلة.

وهذا يعني في المحصلة أن $X \sim Binomial\,(x;\,40,\,0.125)$.

مثال (5.6): في دوري لمباريات كرة القدم كان احتمال أن يفوز الفريق A في أي مباراة يلعبها هو 2/3 . فإذا لعب هذا الفريق 4 مباريات، فأوجد احتمال أن يفوز:

1. بثلاث مباريات 2. بمباراة واحدة على الأقل 3. بأكثر من نصف عدد المباريات التي لعبها.

الحل:

لنفرض أن X هو متغير يمثل صفة النجاح، (وهو الفوز في هذه التجربة)، في عدد $n = 4$ مباريات (محاولات)، والصفة المضادة هي عدم الفوز أي الفشل، واحتمال الفوز (والذي يمثل المعلمة $p = 2/3$) هو ثابت في أي محاولة أو مباراة يلعبها الفريق. إذن نستطيع اعتبار أن المتغير X يتبع توزيع ذي الحدين بالمعالم $n= 4$ و $p = 2/3$ ، وبالتالي نستطيع استخدام التوزيع في إيجاد الاحتمالات المطلوبة[1] في المثال بعد التعويض بقيم المعالم في التوزيع:

$$X \sim Binomial\,(x;\,4,\,2/3)$$

$$P\left(x;4,\frac{2}{3}\right) = P(X = x) = C_x^4\,(\tfrac{2}{3})^x\,(\tfrac{1}{3})^{4-x} \quad , \quad x = 0,1,2,3,4$$

1. $$P\left(\text{فوز الفريق } A \text{ في 3 مباريات}\right) = P(X = 3) = C_3^4\,(\tfrac{2}{3})^3\,(\tfrac{1}{3})^{4-3} = \frac{32}{81} = 0.40$$

2. $$P(X \geq 1) = P(1) + P(2) + P(3) + P(4) = 1 - P(0)$$

$$= 1 - \left[C_0^4\,(\tfrac{2}{3})^0\,(\tfrac{1}{3})^{4-0}\right] = 1 - \frac{1}{81} = \frac{80}{81} = 0.99$$

3. $$P(X > 2) = P(3) + P(4) = C_3^4\left(\tfrac{2}{3}\right)^3\left(\tfrac{1}{3}\right)^{4-3} + C_4^4\left(\tfrac{2}{3}\right)^4\left(\tfrac{1}{3}\right)^{4-4}$$

$$= \frac{32}{81} + \frac{16}{81} = \frac{48}{81} = 0.59$$

تعريف (8.6): **توقع وتباين توزيع ذي الحدين:** إذا كان $X \sim Binomial\,(x;\,n,\,p)$ فإن توقع وتباين المتغير X يعرّف بالصورة:

$$E(X) = n\,p \quad , \quad Var(X) = n\,p\,(1 - p)$$

ملاحظة: يمكن اشتقاق صيغة التوقع والتباين لتوزيع ذي الحدين (تعريف (8.6)) بالصورة التالية:

لدينا $X \sim Binomial\,(x;\,n,\,p)$ ، ومن تعريف (4.6) وتعريف (7.6) نستطيع القول بأن المتغير X هو مجموع عدد النجاحات في الـ n محاولة، وإذا ما عرّفنا متغير المؤشر أو الدليل I (Indicator variable) بأنه يمثل نتيجة كل محاولة من محاولات بيرنوللي، حيث

[1] لاحظ أنه يمكن دائما التعامل مع معطيات التجربة وحساب الاحتمالات بصورة مباشرة، إلا أن استخدام صيغة التوزيع (ذي الحدين في هذه الحالة) سيكون أسرع في الحل.

$$I = \begin{cases} 0 & : \text{فشل} \\ 1 & : \text{نجاح} \end{cases}$$

وعلما بأن كل المحاولات مستقلة، نستطيع كتابة

$$X = I_1 + I_2 + \cdots + I_n$$

وبأخذ التوقع للطرفين

$$E(X) = E(I_1) + E(I_2) + \cdots + E(I_n)$$

وحيث أن $I \sim Bernoulli\ (x; p)$ فإن $E(I_j) = p$ ، و $Var(I_j) = p(1-p)$ حيث $j = 1, \dots, n$ وبالتالي يكون

$$E(X) = \overbrace{p + p + \cdots + p}^{n \text{ من المرات}} = n\,p$$

وبالمثل فإن

$$Var(X) = Var(I_1) + Var(I_2) + \cdots + Var(I_n)$$

$$= \overbrace{p(1-p) + p(1-p) + \cdots + p(1-p)}^{n \text{ من المرات}} = n\,p\,(1-p)$$

مثال (6.6): أوجد توقع وتباين المتغير العشوائي X في المثال (5.6).

الحل:

$$E(X) = n\,p = 4 \times \frac{2}{3} = \frac{8}{3} = 2.67$$

$$Var(X) = n\,p\,(1-p) = 4 \times \frac{2}{3} \times \frac{1}{3} = \frac{8}{9} = 0.89$$

تعريف (9.6): **الدالة المولدة للعزوم لتوزيع ذي الحدين:** إذا كان $X \sim Binomial\ (x; n, p)$ فإن الدالة المولدة للعزوم له تعرّف بالصيغة:

$$\mu_x(t) = E(e^{tX}) = \left((1-p) + pe^t\right)^n$$

ملاحظة: للوصول إلى الصيغة السابقة (تعريف (9.6))، يمكن استخدام مفكوك ذي الحدين[1] بعد التعويض في الصيغة $\mu_x(t) = \sum_x e^{tX} P(x)$.

مثال (7.6): استخدم الدالة المولدة للعزوم لتوزيع ذي الحدين لاشتقاق التوقع والتباين للمتغير العشوائي X .

الحل:

بما أن $X \sim Binomial\ (x; n, p)$ و $\mu_x(t) = E(e^{tX}) = \left((1-p) + pe^t\right)^n$ فإن:

$$\acute{\mu}_1 = \frac{\partial \mu_x(t)}{\partial t} = n\left((1-p) + pe^t\right)^{n-1}(pe^t)\Big|_{t=0} = n\left((1-p) + p\right)^{n-1}p\ = np$$
$$= E(X)$$

---

[1] الصيغة العامة لمفكوك ذي الحدين (Binomial expansion) هي $(a+b)^n = \sum_{i=0}^{n} C_i^n \,.\, b^i . a^{n-i}$.

$$\acute{\mu}_2 = \frac{\partial^2 \mu_x(t)}{\partial t^2} = n(n-1)\left((1-p)+pe^t\right)^{n-2}(pe^t)^2 + n\left((1-p)+pe^t\right)^{n-1}(pe^t)\Big|_{t=0}$$

$$= n(n-1)\left((1-p)+p\right)^{n-2}p^2 + n\left((1-p)+p\right)^{n-1}p = n(n-1)p^2 + np$$
$$= n^2p^2 - np^2 + np = E(X^2)$$

وبالتالي فإن

$$\mu_2 = \acute{\mu}_2 - \acute{\mu}_1^2 = n^2p^2 - np^2 + np - n^2p^2 = np - np^2 = n.p.(1-p) \quad = Var(X)$$

ملاحظة: يمكن استخدام جدول توزيع ذي الحدين (Binomial Table) للحصول على القيم الاحتمالية للتوزيع عندما تكون الحسابات طويلة ومعقدة. ويتم ذلك بعد التعويض بقيمة حجم العينة n (عدد المحاولات)، وقيمة معلمة التوزيع p .إلا أننا لن نتطرق لهذا الجدول باعتبار أن كل البرامج (الحزم) الإحصائية تتضمن طرق إيجاد القيم الاحتمالية للتوزيعات المختلفة.

3.2.6 التوزيع متعدد الحدود (Multinomial Distribution)

يمكن اعتبار التوزيع متعدد الحدود كامتداد أو توسع لمفهوم توزيع ذي الحدين، فهو يمثل الحالة العامة للتجارب العشوائية التي ينتج عنها مجموعة محددة من الحالات. فإذا كان لكل محاولة ضمن التجربة أكثر من نتيجتين، فإن توزيع ذي الحدين يصبح التوزيع متعدد الحدود.

تعريف (10.6): **التوزيع متعدد الحدود:** إذا كانت $E_1, E_2, \ldots, E_k$ هي نتائج أي محاولة في التجربة العشوائية باحتمالات مناظرة $p_1, p_2, \ldots, p_k$ ، حيث $\sum_{i=1}^{k} p_i = 1$ ، $p_i > 0$ لكل $i = 1, 2, \ldots, k$ ، فإن المتغيرات العشوائية $X_1, X_2, \ldots, X_k$ والتي تمثل عدد مرات حدوث هذه النتائج في n محاولة مستقلة يكون لها توزيع احتمالي مشترك يعرف بالتوزيع متعدد الحدود بالمعالم n و $p_1, p_2, \ldots, p_k$ ، وتكون له الصيغة:

$$P(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots p_k, n) = \binom{n}{x_1, x_2, \ldots, x_k} . p_1^{x_1} . p_2^{x_2} \ldots p_k^{x_k}$$

حيث $\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1! x_2! \ldots x_k!}$ و $\sum_{i=1}^{k} x_i = n$.

مثال (8.6): في تجربة إلقاء زهري نرد 6 مرات أوجد احتمال الحصول على مجموع للوجهين يساوي 7 أو 11 مرتين، ورقمين متساويين في الوجهين مرة واحدة، وعدم الحصول على الحدثين السابقين ثلاث مرات.

الحل:

لاحظ أن المطلوب هو حساب احتمال واحد يحتوي على ثلاثة أحداث هي:

الحصول على مجموع للوجهين يساوي 7 أو 11 ويساوي E_1 ، الحصول على رقمين متساويين في الوجهين ويساوي E_2 ، و عدم الحصول على E_1 أو E_2 ويساوي E_3 .

وهذه الأحداث الثلاثة E_1 ، E_2 و E_3 تمثل المتغيرات X_1 ، X_2 ، و X_3 على الترتيب. وبتكرار هذه الأحداث أو النتائج حسب المطلوب تكون قيم المتغيرات الثلاثة هي:

E_1 مرتين $\leftarrow x_1 = 2$ ، E_2 مرة واحدة $\leftarrow x_2 = 1$ ، و E_3 ثلاث مرات $\leftarrow x_3 = 3$

وبما أن فراغ العينة لهذه التجربة هو $S = \{(1, 1), (1, 2), ..., (6, 6)\}$ فإن احتمالات الأحداث المطلوبة تكون

$$P(E_1) = \frac{2}{9} = p_1 \quad , P(E_2) = \frac{1}{6} = p_2 \quad , P(E_3) = \frac{11}{18} = p_3$$

ويكون الاحتمال المطلوب في المثال، حيث تم تكرار التجربة (إلقاء زهري النرد) 6 مرات، هو

$$P\left(2, 1, 3\,; \frac{2}{9}, \frac{1}{6}, \frac{11}{18}, 6\right) = \frac{6!}{2!\,1!\,3!} \cdot \left(\frac{2}{9}\right)^2 \cdot \left(\frac{1}{6}\right)^1 \cdot \left(\frac{11}{18}\right)^3 = 0.113$$

تعريف (11.6): **التوقع والتباين للتوزيع متعدد الحدود:** إذا كانت $X_1, X_2, \ldots , X_k$ هي متغيرات عشوائية متعددة الحدود بالمعالم n و $p_1, p_2, ..., p_k$ ، فإن التوقع والتباين لها يعرف بالصورة:

$$E(X_i) = n\,P_i \quad , Var(X_i) = n\,P_i\,(1 - P_i)\;, \; i = 1, 2, ..., k$$

ويكون التغاير بين أي متغيرين معرف بالصيغة:

$$Cov(X_j , X_l) = -\,n\,P_j\,P_l \quad , \forall\, j \neq l$$

4.2.6 التوزيع الهندسي (Geometric Distribution)

التوزيع الهندسي للمتغير العشوائي X ينتج هو الآخر عن التجارب العشوائية التي تندرج تحت محاولات بيرنوللي أيضا، حيث تنقسم فيه نتيجتي كل محاولة إلى نجاح (باحتمال p) أو فشل (باحتمال $(1 - p)$)، وتكون كل المحاولات مستقلة. إلا أن الاختلاف في حالة التوزيع الهندسي هو أن المتغير العشوائي X لا يمثل عدد النجاحات في المحاولات (كما هو الحال في توزيع ذي الحدين)، بل يمثل عدد المحاولات التي تمت حتى الحصول على أول حالة نجاح في التجربة العشوائية. ولتوضيح مفهوم التوزيع الهندسي لنأخذ المثال الافتراضي التالي:

في تجربة إلقاء عملة معدنية، إذا ما عرفنا أن حالة النجاح هي الحصول على صورة H، فإن المتغير العشوائي X ، والذي سيتبع التوزيع الهندسي، سيمثل عدد المحاولات التي ستتم حتى الحصول على أول صورة، والتي قد نحصل عليها في أول رمية أو الثانية أو الثالثة أو ... وهكذا، وعلى هذا فإن عدد المحاولات يكون غير منتهي، فنظريا قد لا نحصل على أي صورة عندما تكون النتيجة دائما هي كتابة T.

وإذا ما عرّفنا k على أنه عدد المحاولات فإن احتمال الحصول على صورة في هذه التجربة سيكون

$$P_X(k) = P(\overbrace{\text{فشل ، فشل ، ... ، فشل}}^{k-1}\text{ ، نجاح})$$

حيث أن التجربة لن تتوقف حتى نحصل على نجاح، أي نحصل على صورة. وحيث أن المحاولات مستقلة، نستطيع كتابة:

$$P_X(k) = \overbrace{P(\text{فشل})\,.\,P(\text{فشل})\,.\,\ldots\,.\,P(\text{فشل})}^{k-1}\,.\,P(\text{نجاح})$$

$$= \overbrace{(1-p)\,.\,(1-p)\,.\,\ldots\,.\,(1-p)}^{k-1}\,.\,p = (1-p)^{k-1}\,p \quad , k = 1, 2, \ldots$$

وهكذا نستطيع تعريف التوزيع الهندسي بالشكل التالي:

تعريف (12.6): **التوزيع الهندسي** (Geometric distribution): يقال أن المتغير العشوائي X يتبع توزيع هندسي بمعلمة p ، ($X \sim Geometric\ (x; p)$)، إذا كانت دالة الكتلة الاحتمالية له على الصورة:

$$P(x;p) = P(X=x) = p\,(1-p)^{x-1} \quad , x = 1, 2, 3, \ldots \quad , 0 \le p \le 1$$

ملاحظة: يسمى التوزيع الهندسي بهذا الاسم لأن مجموع الاحتمالات فيه يمثل مجموع متوالية هندسية كما هو معرّف في الصورة العامة:

$$\sum_{x=0}^{\infty} r^x = \frac{1}{1-r}$$

حيث $|r| < 1$ هو عدد حقيقي.

مثال (9.6): إذا كان احتمال حدوث عطل بإحدى مضخات المياه في مشروع زراعي هو 0.02 خلال ساعة واحدة. فأوجد احتمال عدم حدوث عطل في أي مضخة في هذا المشروع خلال الساعتين القادمتين.

الحل:

لنفرض أن المتغير X يمثل عدد الفترات أو الوحدات الزمنية (التي مدتها ساعة واحدة) حتى حدوث أول عطل في المضخة، حيث النجاح هنا هو حدوث العطل، فيكون الاحتمال المطلوب هو

$$P(\text{عدم تعطل المضخة خلال ساعتين}) = P(X \ge 3)$$

أي حساب احتمال أن أول عطل لن يحدث في فترة الساعة الأولى أو الثانية بل في فترة الساعة التي تليها وهي الساعة الثالثة، وهكذا فإن

$$P(X \ge 3) = P(X=3) + P(X=4) + \cdots = \sum_{x=3}^{\infty} P(x;p)$$

$$= 1 - P(X<3) = 1 - \sum_{x=1}^{2} P(x;p)$$

وحيث أن

$$P(x;p) = p\,(1-p)^{x-1} = (0.02)(0.98)^{x-1} \ , x = 1,2, \ldots$$

فإن

$$P(X \ge 3) = 1 - [P(1) + P(2)] = 1 - [0.02 + (0.98).(0.02)] = 0.96$$

ملاحظة: إذا كان المطلوب، مثلا، حساب احتمال تعطل المضخة في أي عدد من الفترات الزمنية، (على اعتبار أن كل فترة مدتها ساعة)، فيتم التعويض عن هذا العدد في التوزيع الهندسي فنحصل على الاحتمال المطلوب. فمثلا، إذا كان المطلوب إيجاد احتمال تعطل المضخة في الساعة الرابعة، فيكون

$$P(X=4)=(0.98)^{4-1}.(0.02)=0.02$$

تعريف (13.6): **التوقع والتباين للتوزيع الهندسي:** إذا كان X متغير عشوائي يتبع التوزيع الهندسي بمعلمة p، فإن توقعه وتباينه يعرف بالصورة:

$$E(X)=\frac{1-p}{p} \quad , \quad Var(X)=\frac{1-p}{p^2}$$

مثال (10.6): أوجد التوقع والتباين لعدد الفترات الزمنية التي مدتها ساعة واحدة اللازمة لحدوث عطل في مضخة المياه في المثال السابق (مثال (9.6)).

الحل:

$$E(X)=\frac{1-p}{p}=\frac{1-0.02}{0.02}=49 \text{ فترة} \quad , \quad Var(X)=\frac{1-p}{p^2}=\frac{1-(0.02)}{0.02^2}=2450 \text{ فترة}$$

تعريف (14.6): الدالة المولدة للعزوم للتوزيع الهندسي: إذا كان $X \sim Geometric\ (x;\, p)$ فإن الدالة المولدة للعزوم له تعرف بالصورة:

$$\mu_x(t)=E(e^{tx})=\frac{pe^t}{1-(1-p)e^t}$$

5.2.6 توزيع ذي الحدين السالب (Negative Binomial Distribution)

إن توزيع ذي الحدين السالب، (أو كما يسمى في بعض الكتب **توزيع باسكال** (Pascal))، يجمع بين طبيعتي توزيعين احتماليين هما توزيع ذي الحدين والتوزيع الهندسي. فطبيعة التجارب العشوائية التي يرتبط بها توزيع ذي الحدين السالب أيضا تتكون من مجموعة من المحاولات المستقلة، بحيث ينتج عن كل محاولة نتيجتين متضادتين هما النجاح أو الفشل. وكما شاهدنا سابقا، فإن التوزيع الهندسي يتعامل مع الحالات التي نكون فيها مهتمين بحدوث أول حالة نجاح، والسؤال المطروح هنا هو ماذا إن كنا مهتمين بحدوث ثاني أو ثالث أو رابع أو ... حالة نجاح ؟. لتوضيح الإجابة بشكل عملي نأخذ المثال التوضيحي التالي:

لنفرض أنه في أحد المستشفيات تم اعتبار أن أي حالة ولادة هي تجربة عشوائية ينتج عنها مولودة أنثى (ولتكن هي النجاح)، أو مولود ذكر (ولتكن الفشل)، بإهمال حالات ولادة التوائم. وتم مراقبة نتائج عدة عمليات ولادة خلال أحد الأيام، ونريد إيجاد احتمال الحصول على ثلاث مواليد من الإناث عند الوصول لحالة الولادة (المحاولة) الخامسة. هذا الاحتمال هو عبارة عن احتمال تقاطع الحدثين التاليين:

P (الحصول على 3 إناث (نجاحات) في الولادات (المحاولات) الخمسة الأولى)

$= P$ (الحصول على أنثى في الولادة الخامسة $\cap$ الحصول على 2 إناث في الولادات الأربع الأولى)

$$P(E) = P(E_1 \cap E_2)$$

وحيث أن الولادات (المحاولات) هي مستقلة عن بعضها البعض فإن

$$P(E) = P(E_1) \times P(E_2)$$

ونلاحظ أن الاحتمال $P(E_1)$ يندرج تحت مفهوم توزيع ذي الحدين بالمعالم $n=4$ و $P=1/2$، حيث أن احتمال النجاح والذي يمثل الحصول على مولودة أنثى هو 1/2، وكذلك الاحتمال $P(E_2)$ يساوي احتمال الحصول على مولودة أنثى في أي محاولة، وبالتالي نستطيع الحصول على الاحتمال المطلوب بالصورة التالية:

$$P(E) = \left(C_2^4\ (\tfrac{1}{2})^3\ (\tfrac{1}{2})^{4-2}\right) \times \left(\tfrac{1}{2}\right)$$

وبصورة عامة، إذا كان احتمال النجاح هو p واحتمال الفشل هو $(1-p)$ ، وكان المطلوب هو إيجاد احتمال الحصول على r نجاح (مولودة أنثى طبقا لهذا المثال التوضيحي)، في المحاولات (الولادات) الخمسة الأولى فإن هذا يعني حساب

$$P_X(5) = P(X=5) = (C_{r-1}^{5-1}\, p^{r-1}\, (1-p)^{5-r}) \times p = C_{r-1}^{5-1}\, p^r\, (1-p)^{5-r}$$

وهكذا يمكننا تعريف المتغير العشوائي X ، والذي يمثل عدد المحاولات التي تمت حتى الحصول على النجاح رقم r (النجاح الرائي) كما يلي:

تعريف (15.6): **توزيع ذي الحدين السالب** (Negative Binomial Distribution): يقال أن المتغير العشوائي X يتبع توزيع ذي الحدين السالب بالمعالم (r, p) إذا كانت له دالة الكتلة الاحتمالية التالية:

$$P(x;r,p) = C_{r-1}^{x-1}\, p^r\, (1-p)^{x-r} \quad , \qquad x = r, r+1, r+2, \ldots$$

$$r = 2, 3, \ldots \quad , \quad 0 \le p \le 1$$

ويرمز لذلك بالصيغة $X \sim N.Binomial\,(x; r, p)$.

ملاحظات:

1. إذا ما تم تعريف المتغير العشوائي $Y = X - r$ فإن التوزيع الاحتمالي للمتغير Y يكون:

$$P_Y(y) = C_{r-1}^{y+r-1}\, p^r\, (1-p)^y = C_y^{y+r-1}\, p^r\, (1-p)^y \ , y = 0, 1, 2, \ldots$$

حيث أنه $C_y^{x-1} = C_{r-1}^{x-1}$ لأن $x - 1 = y + r - 1$ ، (راجع خواص التوافيق ضمن طرق العد في البند (1.3.4)).

وحيث أن المتغير Y يمثل عدد حالات الفشل قبل حدوث النجاح رقم r ، والحد الأول في توزيع Y من الدالة الاحتمالية يمكن كتابته بالصورة:

$$C_y^{y+r-1} = (-1)^y . C_y^{-r}$$

بالتالي فإن توزيع المتغير Y يمكن كتابته بالصورة

$$P_Y(y) = C_y^{-r}\, p^r\, (-(1-p))^y \quad , y = 0, 1, 2, \dots \quad , r = 2, 3, \dots$$

وهذا هو السبب في تسمية توزيع ذي الحدين السالب بهذا الاسم.

2. إذا ما تم التعويض بالقيمة $r = 1$ في التعريف (15.6) فإن توزيع ذي الحدين السالب بالمعالم 1 و p يصبح

$$P(x; 1, p) = C_0^{x-1}\, p\, (1-p)^{x-1} = \frac{(x-1)!}{(x-1)!\, 0!}\, p\, (1-p)^{x-1} = p\, (1-p)^{x-1} \quad , x = 1, 2, 3, \dots$$

وهذا يعني أن التوزيع الهندسي يمكن اعتباره حالة خاصة من توزيع ذي الحدين السالب وذلك عندما يكون $r = 1$ بمعنى أن يكون المتغير X هو عدد المحاولات التي تمت حتى الحصول على النجاح الأول.

مثال (11.6): بافتراض أن بعض الدراسات النفطية أثبتت أن احتمال الحصول على الغاز الطبيعي في إحدى المناطق الصحراوية الليبية من أي بئر يتم حفرها عشوائيا هو 0.2 . أوجد احتمال الحصول على الغاز:

1. في المرة الثالثة وذلك عند حفر خمسة آبار.
2. في المرة الثالثة في أقل من خمسة محاولات لحفر الآبار.

الحل:

1. لدينا $p = 0.2$ و $r = 3$ ، فيصبح شكل توزيع ذي الحدين السالب

$$P_X(x) = P(x; 3, 0.2) = C_2^{x-1}\, (0.2)^3\, (0.8)^{x-3} \quad , x = 3, 4, 5, \dots$$

وحيث أن المطلوب هو إيجاد قيمة التوزيع عند حفر البئر الخامسة أي عند $x = 5$ ، حتى الحصول على النجاح (الحصول على الغاز) رقم $r = 3$ فإن

$$P_X(5) = C_2^{5-1}\, (0.2)^3\, (0.8)^{5-3} = 0.03$$

2. المطلوب هو

$$P(X < 5) = \sum_{x=3}^{4} C_2^{x-1}\, (0.2)^3\, (0.8)^{x-3} = P(3) + P(4)$$
$$= C_2^2\, (0.2)^3\, (0.8)^0 + C_2^3\, (0.2)^3\, (0.8)^1 = 0.008 + 0.019$$
$$= 0.027$$

تعريف (16.6): **التوقع والتباين لتوزيع ذي الحدين السالب:** إذا كان X متغير عشوائي يتبع توزيع ذي الحدين السالب بالمعالم (r, p) فإن توقعه وتباينه يعطى بالصورة:

$$E(X) = \frac{r(1-p)}{p} \quad , \quad Var(X) = \frac{r(1-p)}{p^2}$$

مثال (12.6): أوجد التوقع والتباين لعدد الآبار التي يتم من خلالها الحصول على الغاز الطبيعي في ثلاث محاولات باستخدام بيانات المثال (11.6).

الحل:

لدينا $r = 3$ و $p = 0.2$ وبالتالي

$$E(X) = \frac{r(1-p)}{p} = \frac{3(1-0.2)}{0.2} = 12 \text{ بئر}$$

$$Var(X) = \frac{r(1-p)}{p^2} = \frac{3(1-0.2)}{(0.2)^2} = 60 \text{ بئر}$$

تعريف (17.6): الدالة المولدة للعزوم لتوزيع ذي الحدين السالب: إذا كان $X \sim N.Binomial\ (x, r, p)$ فإن الدالة المولدة للعزوم له تعرف بالصورة:

$$\mu_x(t) = E(e^{tx}) = \left(\frac{pe^t}{1-(1-p)e^t}\right)^r$$

6.2.6 التوزيع فوق الهندسي (Hyper-geometric Distribution)

في توزيع ذي الحدين، التوزيع الهندسي، وتوزيع ذي الحدين السالب لاحظنا وجود صفة مشتركة تجمع بينها وهي أن المحاولات في التجربة العشوائية تكون مستقلة عن بعضها البعض، وهذا إن صح التعبير قد يتوافق مع مفهوم اختيار أو سحب العينات بالإرجاع.

في التوزيع فوق الهندسي، لا يزال مفهوم تصنيف التجربة إلى نجاح وفشل قائما إلا أن سحب العينات سيكون بدون إرجاع، وهذا بدوره سينفي صفة الاستقلال عن محاولات التجربة العشوائية.

لنفرض أنه لدينا مجتمع (أو فراغ عينة) يحتوي على N عنصر يمكن تقسيمها إلى r عنصر تصنف على أنها حالات نجاح، و($N - r$) عنصر المتبقية تصنف على أنها حالات الفشل، وتم سحب عينة عشوائية حجمها n عنصر من المجتمع الذي حجمه N بدون إرجاع بحيث تكون $r \leq N$ و $n \leq N$. فإذا ما تم تعريف المتغير العشوائي X بأنه يمثل عدد حالات النجاح في العينة التي حجمها n، فإن X سوف يتبع في هذه الحالة توزيع يعرف بالتوزيع فوق الهندسي.

ولتقريب المفهوم بصورة أفضل لنأخذ المثال التوضيحي التالي؛

في سياق المثال التوضيحي الذي تم تناوله في الجزء (5.2.6)، والخاص بتحديد جنس المولود في أحد المستشفيات، نستطيع إجراء التجربة التالية:

لنفرض أنه في أحد الأيام وجدت $N = 10$ حالات ولادة منها $r = 4$ حالات ولادة لإناث والباقي ذكور، وتم اختيار عينة عشوائية مكونة من $n = 3$ حالات ولادة من الـ N ولادة، فإذا ما تم تعريف المتغير العشوائي X بأنه يمثل عدد الولادات لإناث (حالات النجاح) في العينة n فإن التوزيع الاحتمالي للمتغير X يتم حسابه بالصورة[1] الموضحة أدناه:

أولا: حساب عدد الحالات الممكنة لاختيار x ولادة لإناث من بين $r = 4$ حالة تتمتع بصفة النجاح ويساوي C_x^4.

ثانيا: حساب عدد الحالات الممكنة لاختيار $n - x = 3 - x$ ولادة لغير الإناث من بين $N - r = 10 - 3$ حالة لا تتمتع بصفة النجاح ويساوي C_{3-x}^{10-3}.

ثالثا: حساب عدد الحالات الكلية لاختيار عينة حجمها $n = 3$ من مجتمع حجمه $N = 10$ ويساوي C_3^{10}.

وهكذا، يمكن حساب احتمال الحصول على أي عدد x من الولادات لإناث في العينة $n = 3$ بالصورة

$$P = \frac{C_x^4 \; C_{3-x}^{10-4}}{C_3^{10}} \quad , \quad x = 0, 1, 2, \dots, n \quad , \quad x \leq 4$$

وبصورة عامة، يمكن حساب هذا الاحتمال لأي مجتمع حجمه N وعينة مسحوبة منه حجمها n كما هو موضح في التعريف التالي:

تعريف (18.6): **التوزيع فوق الهندسي**: يقال أن المتغير العشوائي X يتبع التوزيع فوق الهندسي بالمعالم N، n، و r إذا كانت دالة الكتلة الاحتمالية له معرّفة بالصورة:

$$P_X(x) = P(x; N, n, r) = \frac{C_x^r \; C_{n-x}^{N-r}}{C_n^N}$$

حيث

$$x = 0, 1, 2, \dots, n \quad , \quad x \leq r \quad , \quad (n - x) \leq (N - r)$$

ويرمز لذلك بالرمز $X \sim H.\ Geometric\ (x; N, n, r)$.

مثال (13.6): مخزن به 100 جهاز تكييف ياباني الصنع و 200 جهاز تكييف صيني الصنع. تم اختيار أربع أجهزة تكييف بصورة عشوائية بدون إرجاع من المخزن، أوجد احتمال:

1. أن تكون كل الأجهزة التي تم اختيارها يابانية الصنع.
2. أن يوجد جهازين على الأقل في العينة يابانية الصنع.

[1] كما هو الحال عند حساب مسائل الاحتمالات في الفصل الرابع.

الحل:

1. حيث أنه يمكن اعتبار $N = 300$ ، $n = 4$ ، و $r = 100$ إذن يمكن استخدام التوزيع فوق الهندسي بالمعالم:

$$P(x; N, n, r) = P(x; 300, 4, 100) = \frac{C_x^{100}\, C_{4-x}^{300-100}}{C_4^{300}}, \quad x = 0, 1, 2, \dots, n$$

وهكذا يكون

$$P(4) = \frac{C_4^{100}\, C_{4-4}^{300-100}}{C_4^{300}} = 0.011$$

.2

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4)$$

$$= \frac{C_2^{100}\, C_2^{200}}{C_4^{300}} + \frac{C_3^{100}\, C_1^{200}}{C_4^{300}} + \frac{C_4^{100}\, C_0^{200}}{C_4^{300}} = 0.407$$

تعريف (19.6): **التوقع والتباين للتوزيع فوق الهندسي:** إذا كان $X \sim H.\ Geometric\ (x; N, n, r)$ فإن توقع وتباين X يعطى بالصورة:

$$E(X) = n\,p \quad , \quad Var(X) = n\,p\,(1-p)\left(\frac{N-n}{N-1}\right)$$

حيث $p = \frac{r}{N}$ ، و $0 \leq p \leq 1$.

مثال (14.6): أوجد التوقع والتباين للمتغير العشوائي X في المثال (13.6).

الحل:

لدينا $p = \frac{r}{N} = \frac{100}{300} = \frac{1}{3}$ ، وبالتالي يكون

$$E(X) = n\,p = 4 \times \frac{1}{3} = 1.33 \text{ جهاز}$$

$$Var(X) = n\,p\,(1-p)\left(\frac{N-n}{N-1}\right) = 4 \times \frac{1}{3} \times \frac{2}{3} \times \left(\frac{300-4}{300-1}\right) = 0.88 \text{ جهاز}$$

ملاحظة: من الجدير بالذكر أنه لا توجد دالة مولدة للعزوم للتوزيع فوق الهندسي لعدم إمكانية حسابها.

7.2.6 توزيع بواسون (Poisson Distribution)

إن توزيع بواسون يعد من أكثر التوزيعات الاحتمالية استخداما في التطبيقات الفيزيائية التي ترتبط بالزمن، وكذلك في الظواهر التي تحدث خلال فترة زمنية مثل عدد المسافرين الواصلين إلى أحد المطارات، عدد السيارات التي تصل إلى تقاطع ما، توزيع ذرات الغبار في فضاء محدد، ...، وغيرها.

فالتجارب العشوائية التي فيها المتغير العشوائي يمثل عدد النتائج التي تقع خلال فترة زمنية محددة أو في نطاق محدد تكون التجارب التي يعرّف عليها توزيع بواسون. فالتجارب التي تجري خلال دقائق، ساعات، أيام، ...، سنوات، مثل عدد الرسائل الإلكترونية المستلمة في أحد المكاتب خلال ساعة، أو عدد المباريات التي تم إلغاؤها بسبب سوء الأحوال الجوية خلال أحد الأشهر، كلها ترتبط بالزمن. أما مثلا أعداد الجراد الذي يهاجم الهكتار الواحد في أحد الحقول الزراعية أو عدد الأخطاء المطبعية في الصفحة الواحدة في كتاب ما فهي تجارب عشوائية ترتبط بنطاق محدد.

ولتعريف توزيع بواسون، يجدر بنا فهم ما يعرف بعملية بواسون أولا.

تعريف (20.6): **عملية بواسون** (Poisson Process): لنفرض أن تجربة عشوائية تم إجراؤها خلال فترة زمنية محددة أو نطاق محدد، وتم تقسيم هذه الفترة (أو هذا النطاق) إلى فئات جزئية صغيرة جدا، عندها يقال أن هذه التجربة العشوائية تمثل عملية بواسون إذا حققت الشروط التالية:

1. إذا كان عدد النتائج التي تحدث في أي فترة جزئية من الفترة الزمنية (أو النطاق المحدد) مستقلا عن عدد النتائج التي تحدث في الفئات الجزئية الأخرى.

2. إذا كان احتمال حدوث نتيجة واحدة في أي فئة جزئية هو نفسه لكل الفئات الجزئية ويتناسب مع طول الفترة الزمنية (أو النطاق المحدد)، ولا يعتمد على عدد النتائج التي تحدث خارج تلك الفئة الجزئية.

3. إذا كان احتمال حدوث أكثر من نتيجة واحدة في الفئة الجزئية الواحدة صغير جدا بحيث يقترب من الصفر.

ولنأخذ المثال التوضيحي التالي لتبسيط مفهوم عملية بواسون:

لنفترض بأننا مهتمين بإيجاد التوزيع الاحتمالي للمتغير العشوائي X الذي يمثل عدد حوادث السيارات عند تقاطع معين خلال أسبوع من الزمن. ولنعتبر أن هذا الأسبوع (الفترة الزمنية) يمكن تقسيمه إلى فئات جزئية صغيرة (دقائق أو ثواني مثلا) بحيث لا يمكن أن يقع في أي فئة جزئية منها أكثر من حادث سير واحد، ولنفرض أيضا أن احتمال أن يقع حادث سير في أي فئة جزئية هو p ، عندئذ سنرى أنه في هذه التجربة العشوائية يكون:

P (وقوع حادث سير في أي فترة جزئية من الزمن) $= p$

P (عدم وقوع أي حادث سير في أي فترة جزئية) $= 1 - p$

P (وقوع أكثر من حادث سير واحد في أي فترة جزئية) $= 0$

وحيث أن وقوع حوادث السير في هذه الفترات الجزئية من الأسبوع يكون مستقلا من فترة لأخرى، تصبح هذه التجربة العشوائية إحدى عمليات بواسون.

ولاحظ أن مجموع عدد الحوادث يكون له توزيع ذي الحدين بالمعالم n و p ، (حيث يمثل n عدد الفترات الجزئية)، لأن طبيعة التجربة سينتج عنها وقوع حادث سير (نجاح) أو عدم وقوعه (فشل) باحتمال ثابت p لكل فترة جزئية.

تعريف (21.6): **توزيع بواسون:** إذا كان X متغير عشوائي يمثل عدد النتائج التي تحدث خلال فترة زمنية معينة أو ضمن نطاق محدد، فإنه يقال أنه يتبع توزيع بواسون بمعلمة λ إذا كانت دالة الكتلة الاحتمالية له معرّفة بالصورة:

$$P_X(x) = P(x;\ \lambda) = \frac{e^{-\lambda}\,\lambda^x}{x!}\quad,\quad x = 0, 1, 2, \dots\quad, \lambda > 0$$

ويرمز لذلك بالرمز $X \sim Poisson\ (x; \lambda)$.

لاحظنا في المثال التوضيحي الذي يلي التعريف (20.6) والخاص بحوادث السير الواقعة خلال أسبوع أن مجموع الحوادث يتبع توزيع ذي الحدين. وإذا ما افترضنا الآن أن عدد تقسيمات الفترة الزمنية (الأسبوع في مثالنا) أصبح كبيرا جدا (جزء من الثانية مثلا) بحيث أصبح احتمال وقوع حادث واحد في أي فترة جزئية ضئيل جدا، عندئذ سيكون توقع عدد الحوادث في هذا الأسبوع (والذي يساوي $n\,p$) تقريبا ثابتا.

وإذا ما فرضنا أن $n\,p = \lambda$ فإن:

$$Binomial(x; n, p) = C_x^n p^x (1-p)^{n-x} = C_x^n \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

و $x = 0, 1, 2, \dots, n$.

ويمكن ببعض الحسابات الرياضية[1] إثبات أن

$$\lim_{n\to\infty} Binomial(x; n, p) = \frac{e^{-\lambda}\,\lambda^x}{x!} = Poisson(x; \lambda)$$

حيث $x = 0, 1, 2, \dots\quad, \lambda > 0$.

مثال (15.6): في إحدى المدن الصغيرة، تم تطعيم 2000 شخص بالغ ضد مرض الأنفلونزا الموسمية من النوع A . وكان احتمال أن يصاب أي شخص من هؤلاء بآثار جانبية من جراء التطعيم هو 0.001 . فأوجد احتمال أن يصاب بتلك الآثار الجانبية من ضمن الـ 2000 شخص:

1. ثلاثة أشخاص. 2. أكثر من شخصين. 3. عدم إصابة أي شخص.

الحل:

لدينا $n = 2000$ ، $p = 0.001$ حيث تندرج طبيعة هذه التجربة تحت مفهوم توزيع ذي الحدين على اعتبار أن النجاح هو الإصابة بالآثار الجانبية، والفشل عكس ذلك.

إلا أننا نلاحظ أن قيمة n هي كبيرة جدا و قيمة p هي صغيرة جدا، وهذا يعني أن توقع الإصابة بالآثار الجانبية ($n.p = 2000 \times 0.001 = 2$) سيكون ثابتا في هذه المدينة، مع ملاحظة أن طبيعة التجربة تعتمد على نطاق محدد (وهو الأشخاص) وليس على فترة زمنية. وهذا يعني أننا نتعامل مع متغير عشوائي (X) يمثل عدد

[1] يمكن الرجوع للإثبات في (Wackerly, D., (2002), Mathematical Statistics with applications, P.125).

الأشخاص الذين يصابون بآثار جانبية من جراء التطعيم. وبالتالي فإن X يكون له توزيع بواسون بالمعلمة : $\lambda=n.p = 2$

$$P_X(x) = P(x;\ 2) = \frac{e^{-2}\ 2^x}{x!} \quad , \quad x = 0, 1, 2, \dots$$

وهكذا نستطيع حساب الاحتمالات المطلوبة كالتالي:

.1

$$P(X = 3) = \frac{e^{-2}\ 2^3}{3!} = 0.18$$

.2

$$P(X > 2) = P(3) + P(4) + \cdots = 1 - P(X \leq 2) = 1 - [P(0) + P(1) + P(2)]$$

$$= 1 - \left[\frac{e^{-2}\ 2^0}{0!} + \frac{e^{-2}\ 2^1}{1!} + \frac{e^{-2}\ 2^2}{2!}\right]$$

$$= 1 - [0.14 + 0.27 + 0.27] = 0.32$$

.3

$$P(X = 0) = \frac{e^{-2}\ 2^0}{0!} = 0.14$$

تعريف (22.5): **توقع وتباين توزيع بواسون:** إذا كان $X \sim Poisson\ (x;\ \lambda)$ فإن توقع وتباين المتغير العشوائي X يعرّف بالصورة:

$$E(X) = \lambda \quad , \quad Var(X) = \lambda$$

مثال (16.6): إذا علمت أن معدل انتشار البكتيريا في إحدى مزارع البكتيريا هو 9 خلايا في الثانية الواحدة، فما هو احتمال أن تحتوي إحدى المزارع على أكثر من ثلاثة خلايا بكتيرية في الثانية؟ ، كذلك أوجد توقع وتباين عدد خلايا البكتيريا المنتشرة في أي مزرعة.

الحل:

لدينا $\lambda = 9$ حيث يمثل المتغير العشوائي X عدد خلايا البكتيريا المنتشرة في الثانية، ويكون له التوزيع الاحتمالي:

$$P_X(x) = P(x;\ 9) = \frac{e^{-9}\ 9^x}{x!} \quad , \quad x = 0, 1, 2, \dots$$

وبالتالي

$$P(X > 3) = 1 - P(X \leq 3)$$

$$= 1 - [P(0) + P(1) + P(2) + P(3)]$$

$$= 1 - [0.0001 + 0.0011 + 0.005 + 0.014] = 0.98$$

ويكون التوقع والتباين لعدد خلايا البكتيريا هو

$$E(X) = \lambda = 9 \quad , \quad Var(X) = \lambda = 9$$

تعريف (23.6): **الدالة المولدة للعزوم لتوزيع بواسون:** إذا كان $X \sim Poisson\ (x;\lambda)$ فإن الدالة المولدة للعزوم له تعرّف بالصورة:

$$\mu_x(t) = E(e^{tx}) = e^{\lambda(e^t-1)}$$

مثال (17.6): أوجد التوقع والتباين لتوزيع بواسون مستخدما دالته المولدة للعزوم.

الحل:

لدينا من التعريف (23.6)

$$\mu_x(t) = e^{\lambda(e^t-1)}$$

وهكذا فإن توقع X هو

$$\acute{\mu}_1 = \left.\frac{\partial \mu_x(t)}{\partial t}\right|_{t=0} = e^{\lambda(e^0-1)} \times \lambda\, e^0 = \lambda = E(X)$$

و

$$\acute{\mu}_2 = \frac{\partial^2 \mu_x(t)}{\partial t^2} = e^{\lambda(e^t-1)} \times \lambda\, e^t + \lambda\, e^t(\lambda\, e^t\, e^{\lambda(e^t-1)})\Big|_{t=0} = \lambda + \lambda^2$$

فيكون تباين X هو

$$\mu_2 = \acute{\mu}_2 - {\acute{\mu}_1}^2 = \lambda + \lambda^2 - \lambda^2 = \lambda = Var(X)$$

3.6 بعض التوزيعات الاحتمالية المتصلة (Some Continuous Probability Distributions)

في الفصل السابق، تم مناقشة مفهوم التوزيعات الاحتمالية المتصلة وكيف أن المتغيرات العشوائية المتصلة تمثل النتائج التي لا يمكن فصل نقاطها عن بعضها البعض (كما هو الحال في التوزيعات المنفصلة)، بل يتم تعريفها على فترات متصلة، مثل أن نقول أن المتغير X يمثل كمية المطر التي تهطل في مدينة بنغازي في فصل الشتاء، والتي قد تأخذ أي قيمة في الفترة (5 ، 0) بوصة، ولا يمكن عد نقاطها.

وهنالك الكثير من الظواهر والمقاييس التي تتبع متغيراتها العشوائية توزيعات متصلة مثل فترات الزمن، الوزن، الارتفاع، الحجم، ....، وغيرها. وكما هو الحال في الجزء السابق (2.6)، سنقوم في هذا البند بتعريف أهم التوزيعات الاحتمالية المتصلة وكذلك التعرف على طبيعتها وخواصها المميزة.

1.3.6 التوزيع المنتظم المتصل (Continuous Uniform Distribution)

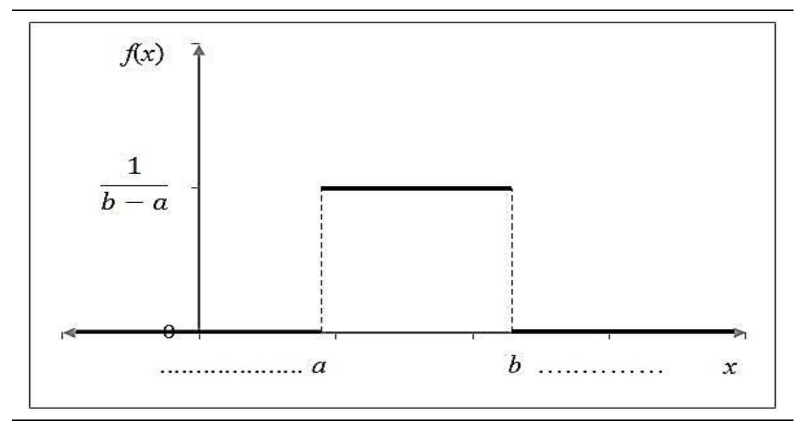
تعرّضنا في البند (1.2.6) للتوزيع المنتظم المنفصل، حيث وضحنا أن المتغير العشوائي في تلك الحالة يأخذ القيم $x_1, x_2, \dots, x_k$ ، $k > 0$ ، وأنه يأخذ قيم احتمالية ثابتة لكل نقطة أو قيمة x_i، $i = 1, 2, \dots, k$. أما الآن، فسيتم عرض الحالة التي يتم فيها تعريف المتغير العشوائي على فترة متصلة.

تعريف (24.6): **التوزيع المنتظم المتصل:** يقال أن X متغير عشوائي يتبع التوزيع المنتظم المتصل على الفترة (a, b) والتي تمثل معالم التوزيع، إذا كانت دالة الكثافة الاحتمالية لـ X معرّفة بالصورة:

$$f_X(x) = f(x; a, b) = \begin{cases} \dfrac{1}{b-a} : a \leq x \leq b \\ 0 \quad : \text{خلاف ذلك} \end{cases}$$

ونستطيع كتابة $X \sim Uniform\ (x;\ a,\ b)$.

والشكل (3.6) يوضح شكل التوزيع المنتظم المتصل.



شكل (3.6): التوزيع المنتظم المتصل على الفترة (a, b).

تعريف (25.6): **التوقع والتباين للتوزيع المنتظم المتصل:** إذا كان $X \sim Uniform\ (x;\ a,\ b)$ فإن توقعه وتباينه يعرّف بالصورة:

$$E(X) = \frac{a+b}{2} \quad , Var(X) = \frac{(b-a)^2}{12} \quad , a \leq b$$

ملاحظة: من الواضح أن توقع وتباين التوزيع المنتظم هو نفسه في حالة كونه منفصل أو متصل، ويمكن الوصول لقيمة التوقع والتباين في التعريف السابق كما يلي:

$$E(X) = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{2(b-a)}[x^2]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b+a)(b-a)}{2(b-a)} = \frac{a+b}{2}$$

و

$$Var(X) = \int_a^b \frac{\left(x - \left(\frac{a+b}{2}\right)\right)^2}{b-a}\,dx = \left[\frac{\left(x - \left(\frac{a+b}{2}\right)\right)^3}{3(b-a)}\right]_a^b = \frac{(b-a)^2}{12}$$

تعريف (26.6): **الدالة المولدة للعزوم للتوزيع المنتظم:** إذا كان $X \sim Uniform\ (x;\ a,\ b)$ فإن الدالة المولدة للعزوم له تعرّف بالصورة:

$$\mu_x(t) = E(e^{tx}) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

مثال (18.6): في أحد المتاجر كان زمن قدوم الزبائن إلى الخزينة يتبع التوزيع المنتظم على الفترة (30 ، 0) دقيقة، حيث أنه خلال أي 30 دقيقة يصل زبون واحد إلى الخزينة. والمطلوب:

1. أوجد احتمال قدوم أي زبون في الخمسة دقائق الأخيرة خلال فترة الـ 30 دقيقة.
2. أوجد التوقع والتباين لهذا التوزيع.

الحل:

1. لدينا $X \sim Uniform\ (x;\ 0,\ 30)$ ، وبالتالي فإن احتمال قدوم أي زبون في الفترة (30 ، 0) دقيقة، هو

$$f(x;0,30) = \frac{1}{30-0} = \frac{1}{30} \quad , \qquad 0 \le x \le 30$$

وهكذا يكون احتمال قدوم أي زبون في الخمسة دقائق الأخيرة يساوي

$$P(25 \le x \le 30) = \int_{25}^{30} \frac{1}{30}\, dx = \frac{30-25}{30} = \frac{1}{6}$$

وهذا منطقي لأن مدة الخمس دقائق تمثل $\frac{1}{6}$ مدة الثلاثين دقيقة.

2. من التعريف (25.6) لدينا

$$E(X) = \frac{a+b}{2} = \frac{0+30}{2} = 15 \text{ دقيقة}$$

أي أن معدل زمن قدوم أي شخص للخزينة هو 15 دقيقة، و

$$Var(X) = \frac{(b-a)^2}{12} = \frac{(0+30)^2}{12} = 75 \text{ دقيقة}$$

ولاحظ أن عدد الزبائن الواصلين إلى الخزينة خلال فترة الـ 30 دقيقة يتبع توزيع بواسون، وهذا يعني أنه يمكن في التجربة العشوائية الواحدة تعريف أكثر من متغير عشوائي بحيث يكون لكل منها طبيعة خاصة (وبالتالي توزيع احتمالي خاص) تختلف باختلاف القيم التي يأخذها، منفصلة مثل عدد الزبائن، أو متصلة مثل زمن الوصول.

2.3.6 **التوزيع الطبيعي** (Normal Distribution)

يعد التوزيع الطبيعي أشهر التوزيعات الاحتمالية على الإطلاق وأكثرها استخداما في علم الإحصاء، خاصة في نظريات التقدير واختبار الفروض. ويمكننا القول[1] بصورة عامة أنه إذا ما تم "تكرار" أي تجربة عشوائية لعدد كبير من المرات فإن المتغير العشوائي الذي يساوي متوسط (أو مجموع) النتائج المتحصل عليها سيؤول[2] توزيعه إلى التوزيع الطبيعي.

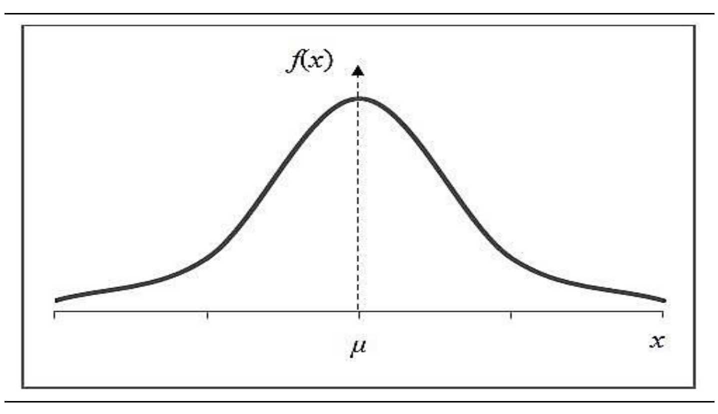
ويعرف التوزيع الطبيعي أيضا **بتوزيع جاوس** (Gaussian Distribution) نسبة إلى العالم الرياضي المعروف[3]، وتشير إليه بعض الكتب بالتوزيع **المعتدل**.

تعريف (27.6): **التوزيع الطبيعي:** يقال أن المتغير العشوائي X يتبع التوزيع الطبيعي بالمعالم μ و σ^2 ، إذا كانت له دالة الكثافة الاحتمالية:

$$f_X(x) = f(x;\ \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\ \sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

حيث $-\infty < x < \infty$ ، $-\infty < \mu < \infty$ ، $\sigma > 0$ ، و $\pi = 3.14$. ويرمز لذلك بالرمز $X \sim N(\mu, \sigma^2)$.

ويأخذ منحنى دالة الكثافة الاحتمالية للتوزيع الطبيعي شكلا مميزا يعرف **بالمنحنى الطبيعي** (Normal Curve)، أو **المنحنى الناقوسي** (Bell-Shaped Curve)، كما هو موضح في الشكل (4.6). حيث يلاحظ من الشكل أن منحنى التوزيع يشبه الناقوس أو الجرس المقلوب، والقيمة μ تقسم المنحنى عموديا إلى قسمين متماثلين، بمعنى أن المنحنى الطبيعي هو منحنى متماثل لتوزيع قيم x ، وأن المنحنى غير مغلق عند الأطراف لأن $-\infty < x < \infty$.



شكل (4.6): منحنى التوزيع الطبيعي.

---

[1] من خلال ما يُعرف **بنظرية النهاية المركزية** والتي سيتم تناولها في الفصل القادم.

[2] يقصد بذلك أن المتغير العشوائي قد يكون له ضمن التجربة توزيع احتمالي معين مختلف عن التوزيع الطبيعي، إلا أنه نتيجة لتكرار التجربة فإن التوزيع الأصلي يؤول للتوزيع الطبيعي في النهاية.

[3] رغم أن العالم **دي موافر** (De Moivere) كان أول من قام باشتقاق المعادلة الخاصة بالمنحنى الطبيعي عام 1733.

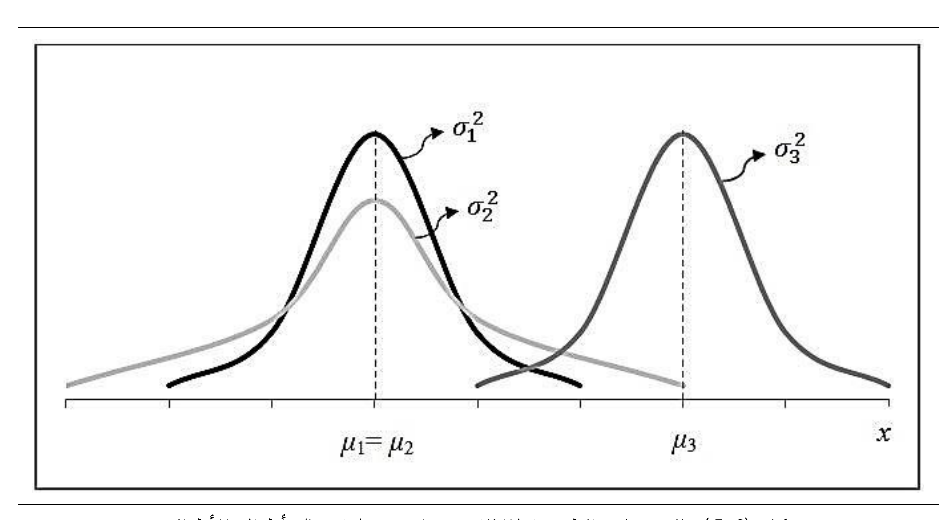
تعريف (28.6): **التوقع والتباين للتوزيع الطبيعي:** إذا كان $X \sim N(\mu, \sigma^2)$ فإن توقع المتغير العشوائي X وتباينه يعرّفان بالصورة:

$$E(X) = \mu \quad , \quad Var(X) = \sigma^2$$

أي أن توقع وتباين التوزيع الطبيعي هما في الواقع معلمتي التوزيع، ويعتمد شكل المنحنى على قيمتهما كما سنرى لاحقا.

ملاحظة: حيث أن التوزيع الطبيعي، كما أسلفنا، هو أكثر التوزيعات أهمية واستخداما، فإنه عادة ما تستخدم الرموز الرومانية μ و σ^2 للتعبير عن التوقع (أو الوسط الحسابي) والتباين، على الترتيب، بصورة عامة في التوزيعات الاحتمالية ككل عند الضرورة.

في ما يلي، سيتم تسليط الضوء على الاختلافات التي تحدث في شكل المنحنى الطبيعي عند تغير قيم معالمه μ و σ^2 . في الشكل (5.6) لدينا ثلاثة توزيعات تمثل أطوال ثلاثة مجموعات من الأطفال في مرحلة الدراسة الابتدائية، ولها المتوسطات μ_1 ، μ_2 ، و μ_3 والتباينات σ_1^2، σ_2^2، و σ_3^2 على الترتيب.



شكل (5.6): المنحنيات الطبيعية لثلاثة توزيعات مختلفة تمثل أطوال الأطفال.

ويلاحظ أن التوزيع الأول والثاني لهما نفس المتوسط إلا أنهما يختلفان في التباين، وهذا قد يعني أن المجموعة الأولى والثانية من الأطفال هم في مرحلة دراسية متقاربة، إلا أن الاختلافات بين أطوال الأطفال في المجموعة الثانية هي أكبر، حيث $\sigma_1^2 < \sigma_2^2$.

من ناحية أخرى، فإن التوزيع الأول، رغم تقارب تباينه من تباين التوزيع الثالث، إلا متوسطه أصغر $\mu_1 < \mu_3$ ، مما يعني أن الأطفال في المجموعة الثالثة هم في مرحلة دراسية متقدمة عن المجموعة الأولى.

ويتمتع المنحنى الطبيعي بالخواص التالية:

1. قيمة المنوال للتوزيع الطبيعي تساوي الوسط الحسابي، لأن أعلى قيمة احتمالية للتوزيع تقع عند $x = \mu$.
2. منحنى التوزيع الطبيعي متماثل حول الوسط الحسابي μ .
3. المنحنى الطبيعي يقترب من المحور الأفقي تقاربيا (أي تقل قيمة الدالة $f_X(x)$) كلما ابتعدت قيمة المتغير العشوائي X عن الوسط μ سواء بالاتجاه الموجب أو الاتجاه السالب.

4. مجموع المساحة تحت منحنى التوزيع الطبيعي يساوي الواحد الصحيح[1].
5. لأي متغير $X \sim N(\mu, \sigma^2)$ يكون[2]:

$$P(\mu - \sigma < X < \mu + \sigma) = 0.68$$

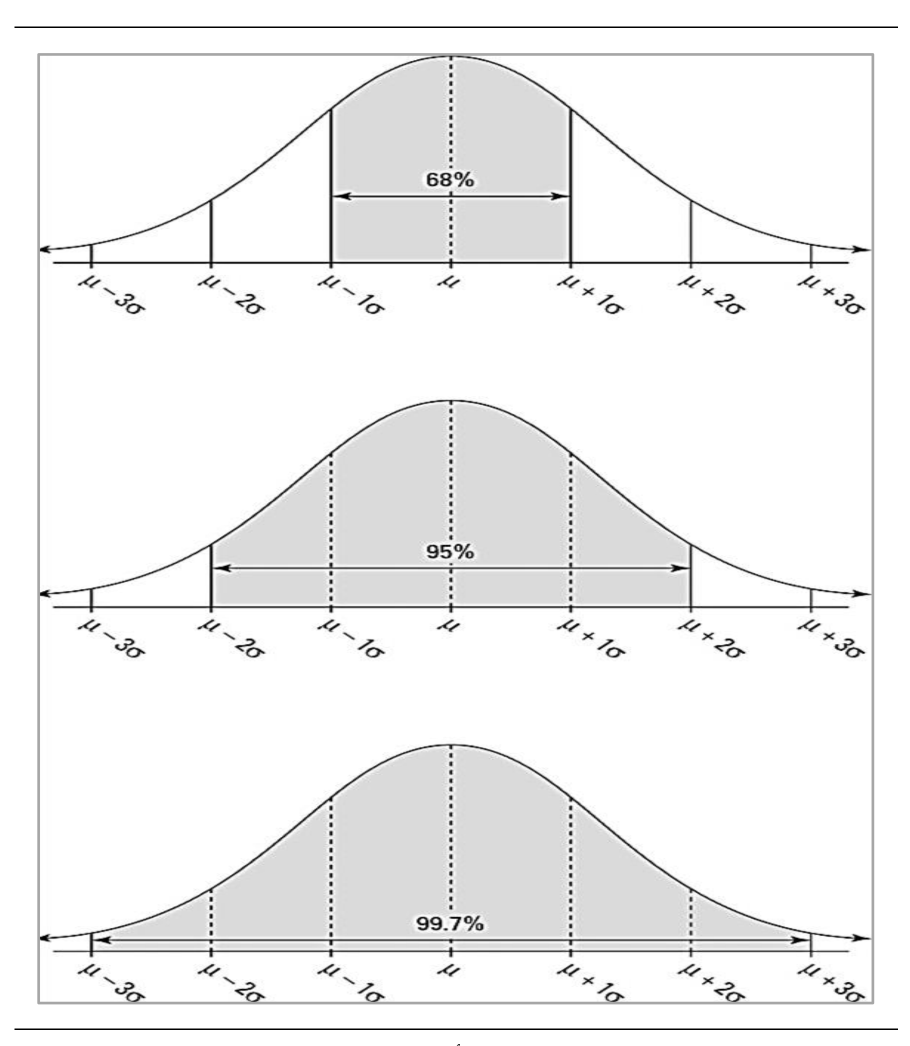
و

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.95$$

و

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.997$$

بمعنى أن 68% ، 95%، و 99.7% من قيم المتغير العشوائي X تقع ضمن الفترات $\mu \pm \sigma$ ، $\mu \pm 2\sigma$ ، و $\mu \pm 3\sigma$ على الترتيب. ويتضح ذلك بيانيا من الشكل[3](6.6).



شكل (6.6): بعض الاحتمالات الأساسية المصاحبة للتوزيع الطبيعي.

[1] حيث أن المساحة الكلية تحت المنحنى الاحتمالي تساوي تكامل الدالة الاحتمالية على كامل نطاق قيم X، والذي يساوي الواحد الصحيح كما هو معروف من خواص التوزيع الاحتمالي.

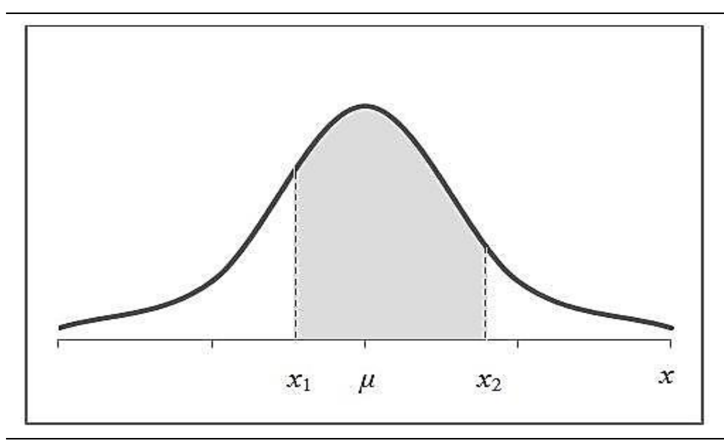
[2] سيتم إثبات هذه الاحتمالات لاحقا في هذا الجزء من الفصل السادس.

[3] المصدر (Rumsey, D., (2003), Statistics for Dummies).

إن التوزيع الطبيعي، كغيره من التوزيعات، يُستخدم لحساب الاحتمالات المختلفة عند قيم المتغير العشوائي X. وكما رأينا سابقا، فإن الاحتمال لأي دالة كثافة احتمالية يتم حسابه باستخدام التكامل للدالة على النطاق المطلوب. فمثلا، إذا كان المطلوب هو حساب احتمال أن تقع قيمة المتغير X بين القيمتين x_1 و x_2 حيث $x_2 > x_1$ ، فإن ذلك يتم بالصورة:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

وهذا يناظر المساحة المظللة تحت المنحنى الطبيعي في الشكل (7.6).



شكل (7.6): المساحة المناظرة للاحتمال $P(x_1 < X < x_2)$.

ومن الناحية العملية، فإننا نفضل دائما الابتعاد عن إجراء التكاملات الرياضية المعقدة والطويلة، لأن الهدف هو حساب الاحتمال وليس إجراء عملية التكامل بحد ذاتها. من ناحية أخرى، فإنه من غير المجدي حساب الاحتمال عن طريق حساب التكاملات لكل قيم X وإعداد جداول تحتوي على تلك القيم واحتمالاتها لأن حساب هذه التكاملات سيعتمد أيضا على قيم معالم التوزيع μ و σ^2 والتي تتغير بتغير التجربة العشوائية أو المجتمع تحت الدراسة، لذلك يتم اللجوء لاستخدام مبدأ تحويل المتغير العشوائي أو ما يعرف **بالمعايرة** (Standardization) "لتثبيت" قيم المعالم. لنعتبر التحويل[1] التالي:

$$Z = \frac{X-\mu}{\sigma}$$

وحيث أن $X \sim N(\mu, \sigma^2)$ ، و $E(X) = \mu$ و $Var(X) = \sigma^2$ فإن توقع وتباين المتغير الجديد Z يتم حسابه بالصورة:

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{E(X)-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$$

و

$$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}Var(X-\mu) = \frac{1}{\sigma^2}Var(X) = \frac{\sigma^2}{\sigma^2} = 1$$

[1] هذا التحويل تم التعرض له سابقا في الفصل الثالث تحت مفهوم الدرجات المعيارية (تعريف (24.3)).

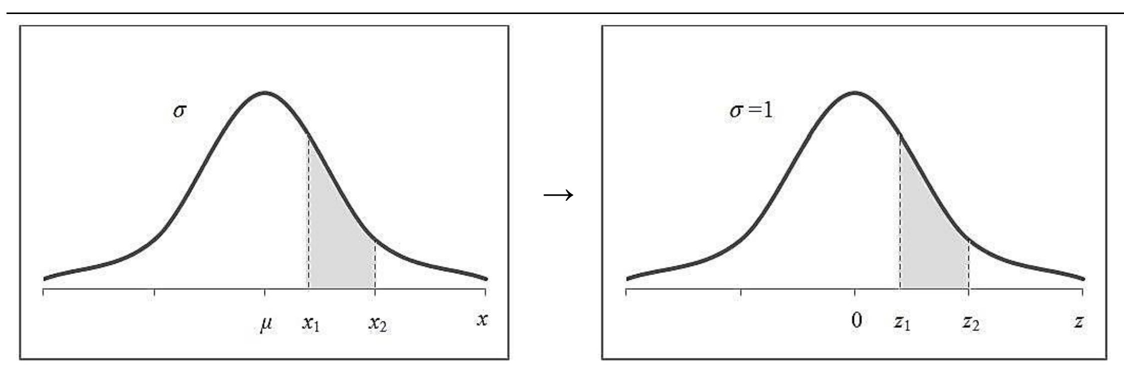
(لأن $Var(\mu) = 0$ حيث أن المتوسط الحسابي هو ثابت) . وهكذا يمكن التعامل الآن مع المتغير العشوائي الجديد Z والذي سيكون له توزيع طبيعي أيضا بالمعالم $\mu = 0$ و $\sigma^2 = 1$.

تعريف (29.6): **التوزيع الطبيعي المعياري** (Standard Normal Distribution): إذا كان X متغير عشوائي يتبع التوزيع الطبيعي بمتوسط μ وتباين σ^2 ، فإن المتغير $Z = \frac{X-\mu}{\sigma}$ يتبع التوزيع الطبيعي بمتوسط يساوي الصفر وتباين يساوي الواحد الصحيح، ويقال أن Z يتبع توزيع طبيعي معياري، وتكون دالة كثافته الاحتمالية بالصورة:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad , \qquad -\infty < z < \infty$$

ويرمز لذلك $Z \sim N(0, 1)$.

وهكذا، فإنه يمكن الآن إيجاد احتمالات وقوع قيم المتغير العشوائي X في أي منطقة عن طريق استخدام التحويل $Z = \frac{X-\mu}{\sigma}$ وإيجاد الاحتمالات المناظرة لقيم Z عوضا عن X . فمثلا إذا كان المطلوب حساب قيمة الاحتمال $P(x_1 < X < x_2)$ فيمكن عندها حساب الاحتمال المناظر $P(z_1 < Z < z_2)$ حيث $z_1 = \frac{x_1-\mu}{\sigma}$ و $z_2 = \frac{x_2-\mu}{\sigma}$ وهذا ما يوضحه الشكل (8.6).



شكل (8.6): الاحتمال $P(x_1 < X < x_2)$ والاحتمال المناظر $P(z_1 < Z < z_2)$ تحت منحنى التوزيع الطبيعي.

مثال (19.6): إذا علمت أن درجات الطلبة في اختبارات أحد مقررات الدراسات العليا تتبع التوزيع الطبيعي بمتوسط 75 درجة وتباين 100 درجة، فأوجد احتمال أن يتحصل أحد الطلبة:

1. على درجة ما بين 80 و 90.
2. على درجة أعلى من 89.5.

الحل:

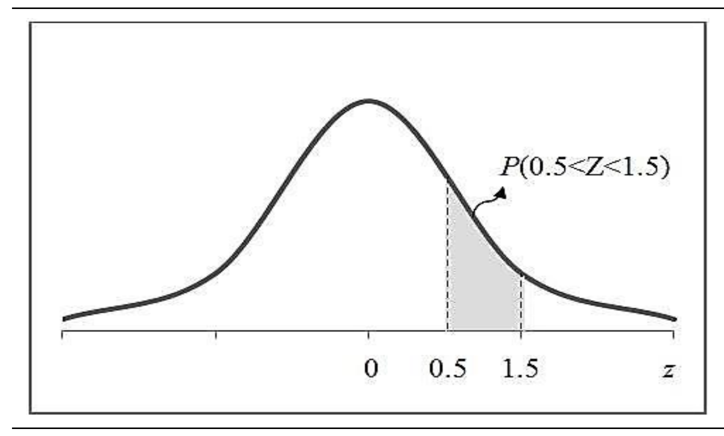
1. المطلوب هو إيجاد الاحتمال $P(x_1 < X < x_2) = P(80 < X < 90)$ ، وحيث أن $\mu = 75$ و $\sigma^2 = 100$، نقوم بتحويل الدرجة X إلى درجة معيارية Z كالتالي

$$Z = \frac{X-\mu}{\sigma} = \frac{X-75}{10}$$

وبالتالي فإن الاحتمال المطلوب يمكن حسابه بالصورة

$$P(80 < X < 90) = P\left(\frac{80-75}{10} < Z < \frac{90-75}{10}\right) = P(0.5 < Z < 1.5)$$
$$= P(Z < 1.5) - P(Z < 0.5)$$

والشكل (9.6) يوضح بيانيا الاحتمال المطلوب حسابه.



شكل (9.6): الاحتمال المطلوب في المثال (19.6) كمساحة تحت المنحنى.

من جدول التوزيع الطبيعي المعياري[1] (ملحق 1، جدول (م2))، ابدأ البحث عن القيمة 1.5 في العامود الأول من اليسار، ثم تحرك أفقيا إلى اليمين إلى العامود تحت القيمة 0.00 ، فتجد أن القيمة هي 0.9332 . فيكون $P(Z < 1.5) = 0.9332$. وبالمثل ابحث من جديد عن القيمة 0.5 في العامود الأول من اليسار ثم تحرك أفقيا إلى اليمين إلى العامود تحت القيمة 0.00 فتجد أن $P(Z < 0.5) = 0.6915$. وهكذا يكون

$$P(80 < X < 90) = 0.9332 - 0.6915 = 0.242$$

2. المطلوب إيجاد

$$P(X > 89.5) = P\left(\frac{X-\mu}{\sigma} > \frac{89.5-75}{10}\right) = P(Z > 1.45)$$
$$= 1 - P(Z \le 1.45)$$

لأن المساحة (الاحتمال) تحت منحنى التوزيع الطبيعي المعياري يتم حسابها بصورة تراكمية إلى اليسار (تدل عليها إشارة أقل من) في جدول التوزيع الطبيعي المعياري.

نبدأ بالبحث في الجدول عن القيمة المطلوبة وهي 1.45 في العامود الأول فنلاحظ أنها غير موجودة بالضبط ولكن أقرب قيمة لها هي 1.4 ، فنتحرك عند هذه القيمة أفقيا حتى نصل إلى العامود تحت القيمة 0.05 وهي القيمة المكملة للقيمة 1.4، $(1.4 + 0.05 = 1.45)$، فنجد أن $P(Z < 1.45) = 0.9265$ وبالتالي يكون

$$P(X > 89.5) = 1 - 0.9265 = 0.0735$$

وهذا يعني أن احتمال أن يتحصل أحد الطلبة على درجة تفوق 89.5 هو احتمال ضئيل جدا. أو يمكن القول بأن نسبة الطلبة الذين قد يحصلوا على درجات أعلى من 89.5 هي 7% تقريبا.

[1] والذي يسمى اختصارا بجدول Z .

ملاحظة: في بعض المسائل قد تكون قيمة Z الناتجة عن التحويل سالبة (وهذا يحدث عندما تكون قيمة المتغير X أقل من المتوسط الحسابي μ)، في هذه الحالة، يتم إيجاد قيمة الاحتمال بالصورة التالية:

لنفرض مثلا أننا نريد إيجاد الاحتمال $P(Z \leq -1.45)$، نبحث أولا في العامود الأول من جدول Z عن القيمة الأقرب للقيمة 1.45- فنجد أنها 1.4- ، ثم نتحرك أفقيا حتى نصل للعامود تحت القيمة 0.05 وهي القيمة المكملة للقيمة 1.4 (بإهمال الإشارة السالبة). فيكون
$P(Z \leq -1.45) = 0.0735$ ، وهذا يعني أن $P(Z \leq -1.45) = 1 - P(Z \leq 1.45)$.

مثال (20.6): إذا علمت أن المتغير العشوائي X والذي يمثل عدد ساعات العمل الأسبوعية لمندوبي المبيعات في إحدى الشركات يتبع توزيعا طبيعيا بمتوسط 40 ساعة وتباين 36 ساعة. فأوجد قيمة X التي يكون:

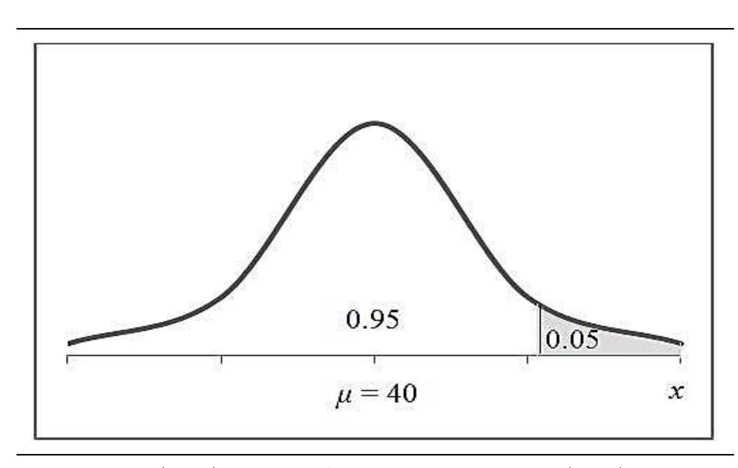
1. 38% من ساعات العمل أقل منها.
2. 5% من ساعات العمل أكثر منها.

الحل:
لاحظ أنه في هذا المثال المعطى هو الاحتمال (ممثلا بالنسبة 38% = 0.38) والمطلوب هو إيجاد قيمة المتغير X . لذلك يجب إيجاد قيمة Z أولا من الاحتمال المعطى ثم إيجاد قيمة X من علاقاتها بـ Z .

1. من جدول Z نبدأ بالبحث عن الاحتمال 0.38 بداخل الجدول (القيم الاحتمالية) فنجد أن أقرب قيمة لها هي 0.3783 ، وهذا الاحتمال يناظر أفقيا القيمة 0.3- وعموديا القيمة 0.01 ، وهكذا يكون $P(Z < -0.31) =$ 0.38 ، أي أن *قيمة* Z المطلوبة هي 0.31- . وبالتالي يتم إيجاد قيمة X بالصورة

$$Z = \frac{X - \mu}{\sigma} \rightarrow X = \sigma Z + \mu = 6 \times (-0.31) + 40 = 38.14 \text{ ساعة}$$

2. نبدأ بالبحث عن الاحتمال المكمل للنسبة 5% (وهو 95%) لأن قيمة X التي يكون 5% من ساعات العمل أكثر منها تساوي الواحد الصحيح مطروحا منه 95% من ساعات العمل كما يوضح الشكل (10.6).



شكل (10.6): المساحة تحت المنحنى الطبيعي للمثال (20.6) الفقرة 2.

ومن جدول التوزيع الطبيعي المعياري، يلاحظ وجود احتمالين كلاهما الأقرب للقيمة 0.95 وهما 0.9495 (تحت العامود 0.04)، و 0.9505 (تحت العامود 0.05)، وكلاهما يناظر القيمة 1.6 أفقيا. في هذه الحالة يتم أخذ

الوسط الحسابي للقيمتين 0.05 و 0.04 ثم يجمع هذا الوسط مع القيمة 1.6 فنحصل على $1.645 = 1.6 + \left(\frac{0.04+0.05}{2}\right)$ ، وهكذا يكون $P(Z < 1.645) = 0.95$ ، وبالتالي: $X = 6 \times 1.645 + 40 = 49.87$ ساعة

ملاحظة: إذا وضعنا $x_1 = \mu - \sigma$ و $x_2 = \mu + \sigma$ يمكننا كتابة:

$$P(x_1 < X < x_2) = P(\mu - \sigma < X < \mu + \sigma) = P\left(\frac{\mu - \sigma - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{\mu + \sigma - \mu}{\sigma}\right)$$
$$= P(-1 < Z < 1) = P(Z < 1) - P(Z < -1) = 0.8413 - 0.1587$$
$$= 0.6836 \cong 0.68$$

وهذا توضيح للفقرة رقم (5) من خواص التوزيع الطبيعي التي تم ذكرها سابقا، وبالمثل يكون الحال عند التعويض بالقيم $\mu \pm 2\sigma$ و $\mu \pm 3\sigma$.

مثال (21.5): إذا كان $X \sim N(74, 49)$ فأوجد العشير السادس لتوزيع المتغير X .

الحل:

العشير السادس (D_6) لتوزيع X هو قيمة X التي تقسم المفردات (القيم التي يأخذها X) إلى قسمين بحيث يكون 60% من القيم أقل من قيمة D_6 ، وبالتالي فإنه من جدول Z نجد أن

$P(Z < 0.25) = 0.60$ ، وبالتالي يكون $D_6 = X = 7\times(0.25)+74 = 75.75$.

نظرية (1.6): **تقريب توزيع ذي الحدين إلى التوزيع الطبيعي:**

إذا كان X متغير عشوائي يتبع توزيع ذي الحدين بالمعالم n و p ، بمتوسط $\mu = n\,p$ و تباين $\sigma^2 = n\,p\,(1-p)$ ، فإنه كلما زادت قيمة n بشكل كبير $(n \to \infty)$ فإن المتغير Z يتبع التوزيع الطبيعي المعياري[1]، حيث

$$Z = \frac{X - n\,p}{\sqrt{n\,p(1-p)}}$$

وتعتبر هذه النظرية ذات فائدة كبيرة عند التعامل مع المسائل أو الظواهر التي تتبع تجارب ذي الحدين ويكون فيها عدد المحاولات أو التكرارات n كبير جدا. ويوجد أيضا تقريب لتوزيع بواسون إلى التوزيع الطبيعي كما هو مبين في النظرية التالية:

نظرية (2.6): **تقريب توزيع بواسون إلى التوزيع الطبيعي:** إذا كان $X \sim Poisson(x; \lambda)$ ، بمتوسط $\mu = \lambda$ وتباين $\sigma^2 = \lambda$ ، فإن المتغير العشوائي

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} \quad , \lambda > 5$$

يتبع التوزيع الطبيعي المعياري.

---

[1] بعض الكتب تضيف إلى النظرية (1.6) القيد $n\,p > 5$ و $n\,(1-p) > 5$.

مثال (22.6): إذا كان احتمال أن يتحصل أي موظف في إحدى المؤسسات على علاوة تشجيعية هو 0.6 . وكان في هذه المؤسسة 100 موظف، فما هو احتمال أن يتحصل أقل من نصفهم على علاوة تشجيعية؟.

الحل:

من الواضح أن هذه التجربة العشوائية تندرج تحت مفهوم توزيع ذي الحدين حيث

$X \sim Binomial\ (x;\ 100,\ 0.6)$، بمتوسط 60 موظف $= 100\times0.6 = \mu$ وتباين 24 موظف $= \sigma^2$ $= 0.4\times0.6\times100$، وحيث أن حجم n كبير فإنه من النظرية (1.6) نستطيع استخدام التقريب إلى التوزيع الطبيعي المعياري.

حيث أن المطلوب هو حساب احتمال أن يتحصل أقل من نصف عدد الموظفين على علاوة، فإن ذلك يعني حساب $P(X < 50)$. إلا أنه يجب مراعاة أن X هو متغير عشوائي منفصل، وأقل من النصف يعني 49 موظف فأقل، وإذا ما تم الأخذ بالاعتبار أن المتغير العشوائي الذي يتبع التوزيع الطبيعي هو متغير متصل فإنه يكون من غير المناسب استخدام القيمة $x = 49$ لأن هذا يكافئ إيجاد احتمال Z لقيمة أقل من 49، وبالتالي نعتبر أن $x = 49.5$ فيكون

$$Z = \frac{49.5 - 60}{4.9} = -2.14$$

ومن جدول Z نحصل على $P(X < 50) \cong P(Z < -2.14) = 0.0162$.

تعريف (30.6): **الدالة المولدة للعزوم للتوزيع الطبيعي:** إذا كان $X \sim N(\mu, \sigma^2)$ فإن الدالة المولدة للعزوم للمتغير X تعرّف بالصورة:

$$\mu_x(t) = E(e^{tx}) = e^{\left(\mu t + \frac{t^2\sigma^2}{2}\right)}$$

ملاحظة: كحالة خاصة، إذا كان $Z \sim N(0, 1)$ فإن الدالة المولدة للعزوم له تعرّف بالصورة:

$$\mu_x(t) = E(e^{tx}) = e^{\left(\frac{t^2}{2}\right)}$$

3.3.6 توزيع جاما (Gamma Distribution)

لاحظنا في التوزيع المنتظم المتصل والتوزيع الطبيعي أن قيم المتغير العشوائي تكون معرفة على فئة الأعداد الحقيقية الموجبة والسالبة، أما في توزيع جاما فإن قيم المتغير تكون غير سالبة (موجبة وتشمل الصفر). وهذا التوزيع له استخداماته المتعددة وخاصة في قياس الزمن. فمثلا الزمن المستغرق للوصول إلى الخزينة في مصرف أو مركز تجاري يتبع عادة توزيع جاما لأنه من البديهي أن لا يتناقص بل هو في زيادة مستمرة. وكذلك الزمن المستغرق لصيانة سيارة أو آلة صناعية، وأيضا الفترة الزمنية المستغرقة لحدوث أعطال في خطوط إنتاج المصانع، وغيرها.

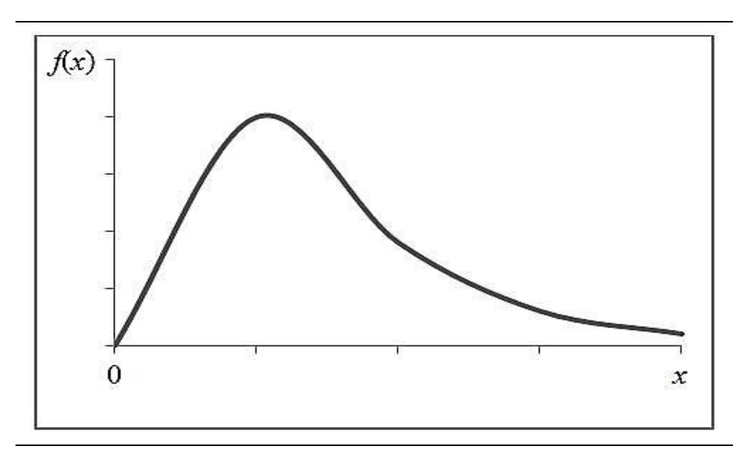
تعريف (31.6): **توزيع جاما:** يقال أن المتغير العشوائي X يتبع توزيع جاما بالمعالم $\alpha > 0$ و $\beta > 0$ إذا كانت دالة الكثافة الاحتمالية له معرّفة بالصورة:

$$f_X(x) = f(x;\alpha,\beta) = \frac{x^{\alpha-1}\, e^{-\frac{x}{\beta}}}{\beta^{\alpha}\,\Gamma(\alpha)} \quad , \quad 0 \le x < \infty$$

ويرمز لذلك بالرمز $X \sim Gamma\ (x;\ \alpha,\ \beta)$. حيث $\Gamma(\alpha)$ تعرف **بدالة جاما** (Gamma Function)، وتأخذ الصيغة:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}\, e^{-x}\, dx$$

والشكل (11.6) يوضح منحنى توزيع جاما (عند قيمة محددة لـ α و β).



شكل (11.6): منحنى توزيع جاما.

ملاحظة: يمكن حساب قيمة دالة جاما عند أي قيمة لـ $\alpha > 0$ بالصورة:

$$\Gamma(\alpha) = (\alpha - 1)!$$

وكذلك فإن

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$

و

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

تعريف (32.6): **التوقع والتباين لتوزيع جاما:** إذا كان $X \sim Gamma\ (x;\ \alpha,\ \beta)$ فإن التوقع والتباين لتوزيع المتغير X يعطى بالصورة:

$$E(X) = \alpha\ .\ \beta \quad , \quad Var(X) = \alpha\ .\ \beta^2$$

حيث $\alpha > 0$ و $\beta > 0$.

مثال (23.6): أحد الموظفين في لجنة الامتحانات النهائية كان يتوقع خروج طالب من اللجنة كل نصف ساعة. فما هو احتمال أن ينتظر هذا المراقب من ساعتين إلى أربع ساعات قبل خروج أربع طلبة؟، على افتراض أن الامتحانات تجري بفترات متلاحقة.

الحل:

لدينا $E(X) = \alpha\,\beta \;\leftarrow\; 1 = \frac{1}{2} \times \beta \leftarrow \beta = 2$ ، حيث أن القيمة المتوقعة لخروج الطلبة هي طالب واحد كل $\alpha = \frac{1}{2}$ ساعة، فيكون معدل خروج الطلبة في الساعة الواحدة هو طالبين. وهكذا فإن $\beta = 2$ سيمثل معدل خروج الطلبة في الساعة الواحدة من الامتحان، و $\alpha = 4$ سيمثل عدد الطلبة المطلوب حساب احتمال خروجهم من الامتحان خلال الفترة الزمنية من ساعتين إلى أربع ساعات. والاحتمال المطلوب هو

$$P(2 \le X \le 4) = \int_2^4 \frac{x^{4-1}\, e^{-\frac{x}{2}}}{2^4\, \Gamma(4)}\, dx$$

وعمليا، يمكن حساب هذا الاحتمال (أو أي احتمال يتضمن حسابات طويلة أو معقدة) باستخدام أي من الحزم الإحصائية، (برنامج اكسل على سبيل المثال)، فيكون

$$P(2 \le X \le 4) = P(X \le 4) - P(X \le 2) = 0.143 - 0.019 = 0.124 \,.$$

تعريف (33.6): **الدالة المولدة للعزوم لتوزيع جاما:** إذا كان $X \sim Gamma\,(x;\, \alpha, \beta)$ فإن الدالة المولدة للعزوم للمتغير X تعرف بالصورة:

$$\mu_x(t) = E(e^{tx}) = (1 - \beta t)^{-\alpha}$$

حيث $\alpha > 0$ و $\beta > 0$.

4.3.6 توزيع بيتا (Beta Distribution)

يُعد توزيع بيتا، هو الآخر، من التوزيعات الغير سالبة، ويكون فيه المتغير العشوائي X معرفا على الفترة $0 \le x \le 1$ ، ولذلك يكون مناسبا للتعامل مع النسب الخاصة بالتجارب العشوائية مثل نسبة الشوائب في مياه الشرب أو نسبة الزمن المستغرق لتصليح آلة كهربائية.

تعريف (34.6): **توزيع بيتا:** يقال أن المتغير العشوائي X يتبع توزيع بيتا بالمعالم $\alpha > 0$ و $\beta > 0$ إذا كانت دالة الكثافة الاحتمالية له معرّفة بالصورة:

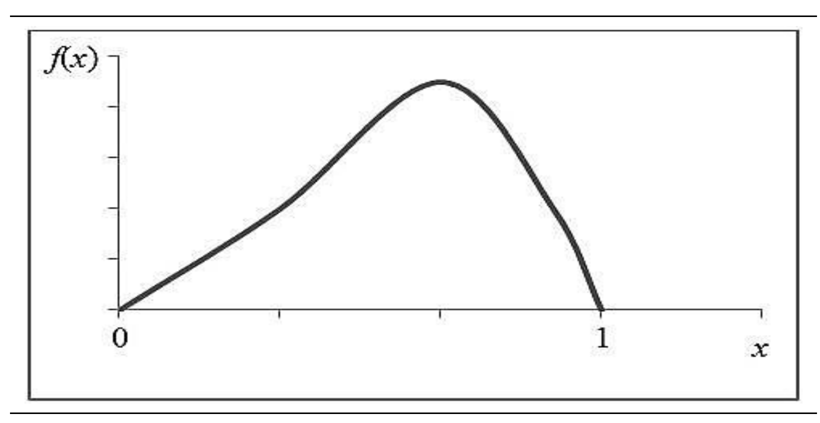
$$f_X(x) = f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\,\beta-1}}{B(\alpha,\beta)} \quad , \quad 0 \le x \le 1$$

ويرمز لذلك بالرمز $X \sim Beta\,(x;\, \alpha, \beta)$. حيث $B(\alpha, \beta)$ هي **دالة بيتا** وتعرّف بالصورة:

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}\,(1-x)^{\beta-1}\,dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

حيث $\Gamma(\cdot)$ هي دالة جاما.

ويأخذ منحنى توزيع بيتا الشكل التالي، (شكل (12.6))، وذلك بحسب قيم $\alpha > 0$ و $\beta > 0$ المختلفة.



شكل (12.6): منحنى توزيع بيتا.

تعريف (35.6): **التوقع والتباين لتوزيع بيتا:** إذا كان $X \sim Beta\ (x; \alpha, \beta)$ فإن توقع وتباين المتغير X يعرّف بالصورة:

$$E(X) = \frac{\alpha}{\alpha + \beta} \quad , \quad Var(X) = \frac{\alpha.\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}$$

حيث $\alpha > 0$ و $\beta > 0$.

ملاحظة: لا توجد دالة مولدة للعزوم لتوزيع بيتا.

مثال (37.6): إذا كان X متغير عشوائي يمثل نسبة كمية المياه التي يتم ضخها إلى مجمع سكني في إحدى الضواحي كل أسبوع، وكان يتبع توزيع بيتا بالمعالم $\alpha = 4$ و $\beta = 2$ ، فأوجد احتمال أن يتم ضخ 90% على الأقل من كمية المياه خلال أحد الأسابيع. ثم أوجد توقع وتباين المتغير X .

الحل:

بناء على قيم معالم توزيع بيتا يكون

$$f(x;4{,}2) = \frac{x^{4-1}(1-x)^{\,2-1}}{B(4{,}2)} = \frac{x^3(1-x)}{\frac{\Gamma(4)\Gamma(2)}{\Gamma(4+2)}} = 20(x^3 - x^4)\ ,\ \ 0 \le x \le 1$$

والمطلوب هو

$$P(X \ge 0.90) = \int_{0.90}^{1} 20(x^3 - x^4)\, dx = 20\left(\left[\frac{x^4}{4}\right]_{0.90}^{1} - \left[\frac{x^5}{5}\right]_{0.90}^{1}\right) = 0.08$$

ويكون توقع وتباين X

$$E(X) = \frac{\alpha}{\alpha+\beta} = \frac{4}{6} = 0.67$$

$$Var(X) = \frac{\alpha\,\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)} = \frac{8}{36 \times 7} = 0.03$$

5.3.6 التوزيع الأسي (Exponential Distribution)

يعد التوزيع الأسي من التوزيعات الهامة في التعامل مع الأعمار الاستهلاكية (كمدة زمنية) للآلات والأدوات الكهربائية والإلكترونية، حيث أن التوزيع يتعامل مع التجارب التي تندرج تحت عمليات بواسون. وفي نفس الوقت، فإن التوزيع الأسي يمكن اعتباره حالة خاصة من توزيع جاما وذلك عندما تكون معلمة جاما $\alpha = 1$.

تعريف (38.6): **التوزيع الأسي:** يقال أن المتغير العشوائي X يتبع توزيعا أسيا بالمعلمة $\beta > 0$ إذا كان له دالة الكثافة الاحتمالية المعرّفة بالصورة:

$$f_X(x) = f(x;\beta) = \frac{1}{\beta}\, e^{-\frac{x}{\beta}} \;\; , \;\; 0 \le x < \infty$$

ويرمز لذلك بالرمز $X \sim Exponential\,(x;\beta)$.

تعريف (39.6): **التوقع والتباين للتوزيع الأسي:** إذا كان $X \sim Exponential\,(x;\beta)$ فإن المتغير X يكون له التوقع والتباين:

$$E(X) = \beta \quad , \quad Var(X) = \beta^2$$

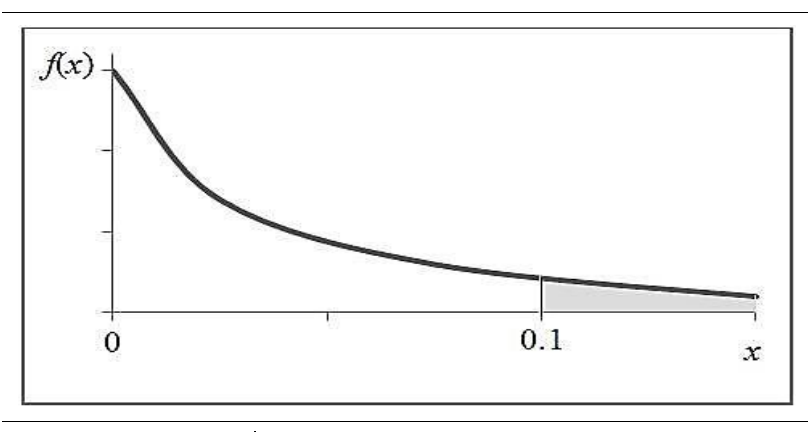
مثال (23.6): في إحدى الشركات المزودة لخدمات الانترنت في ليبيا تم تعريف دخول المشتركين على موقع الشركة بأنه يعتبر عملية بواسون بمتوسط 25 دخول في الساعة الواحدة. أوجد:

1. احتمال أن لا يتم أي دخول إلى موقع الشركة خلال فترة زمنية طولها 6 دقائق.
2. التوقع والتباين للزمن حتى دخول أول مشترك على الموقع.

الحل:

1. لنعتبر أن X هو متغير عشوائي يمثل الزمن (بالساعات) من بداية الفترة الزمنية إلى أول دخول للموقع، في هذه الحالة يكون للمتغير X توزيعا أسيا بمعلمة $\beta = 25$.

الآن نحن مهتمون بإيجاد احتمال أن يأخذ X قيمة تفوق الستة دقائق، (كما يوضح الشكل (13.6))، وحيث أنه يتم قياس وحدات X بالساعات فتكون 6 دقائق = 0.1 ساعة.



شكل (13.6): الاحتمال المطلوب تحت منحنى التوزيع الأسي للمثال (23.6).

وهكذا فإن

$$P(X > 0.1) = \int_{0.1}^{\infty} \frac{1}{25} e^{-\frac{25}{x}} dx = 0.996$$

2. التوقع والتباين للمتغير X هو

$$E(X) = \beta = 25 \quad , \quad Var(X) = \beta^2 = 625$$

تعريف (40.6): **الدالة المولدة للعزوم للتوزيع الأسي:** إذا كان $X \sim Exponential\ (x; \beta)$ فإن الدالة المولدة للعزوم للمتغير X تعرّف بالصورة:

$$\mu_x(t) = E(e^{tx}) = (1 - \beta t)^{-1}$$

وفي نهاية هذا الفصل، تجدر الإشارة إلى وجود بعض التوزيعات الاحتمالية الهامة التي لم نتطرق لها أو لعلاقاتها مع ما تمت دراسته من توزيعات في هذا الفصل، لأن البعض منها يندرج تحت مفهوم يعرف في علم الإحصاء "بتوزيعات المعاينة"، وهو موضوع الفصل القادم.

4.6 تمارين الفصل السادس

تمرين (1.6): إذا كان احتمال أن يتحصل أي موظف في إحدى الشركات الحكومية على إحدى التقييمات (سيئ، مقبول، جيد، جيد جدا، ممتاز) كتقدير لأدائه الوظيفي للعام السابق هو احتمال ثابت، فأوجد:

1. التوزيع الاحتمالي للمتغير العشوائي X الذي يمثل نتيجة تقييم الموظف.
2. التوقع والتباين للمتغير X.
3. التوزيع الاحتمالي التراكمي للمتغير X.

تمرين (2.6): إحدى الدراسات الطبية أثبتت أن 45% من النساء المتزوجات في الوطن العربي ما بين سن 40 و 60 سنة يعانين من الوزن الزائد. فإذا تم اختيار 20 سيدة عربية ضمن تلك الفترة العمرية عشوائيا فأوجد:

1. التوزيع الاحتمالي لعدد النساء اللواتي يعانين من الوزن الزائد ضمن العينة.
2. احتمال أن تعاني نصف النساء في العينة من الوزن الزائد.
3. احتمال أن تعاني امرأتين على الأقل من الوزن الزائد.
4. معدل عدد النساء اللواتي يعانين من الوزن الزائد في العينة وكذلك الانحراف المعياري لعددهن.

تمرين (3.6): قامت إحدى شركات المقاولات بتسليم مشروع سكني يتكون من 500 وحدة سكنية، واتضح فيما بعد أن 25 وحدة من تلك الوحدات به عيوب في أعمال السباكة. فإذا تم اختبار 7 وحدات سكنية من المشروع فأوجد:

1. التوزيع الاحتمالي لعدد الوحدات السكنية ذات العيوب في العينة المختبرة.
2. احتمال وجود ثلاث وحدات بها عيوب.
3. احتمال وجود ثلاث وحدات على الأكثر بها عيوب.
4. احتمال عدم وجود أي وحدة بها عيوب في العينة المختبرة.
5. معدل وتباين عدد العيوب في الوحدات السكنية.

تمرين (4.6): إذا علمت أن نسب حوادث المرور التي وقعت من سنة 2000 إلى سنة 2009 هي موضحة في الجدول التالي فأوجد احتمال أن تحتوي عينة عشوائية مكونة من خمسة حوادث مختارة من تلك السنوات على حادث واحد وقع في الفترة (2000-2001)، حادثين في الفترة (2002-2003)، و حادثين في الفترة ما بين (2006-2007).

نسبة الحوادث	السنة
%18	2000-2001
%23	2002-2003
%16	2004-2005
%27	2006-2007
%16	2008-2009

تمرين (5.6): إذا كان احتمال هبوط طائرة في إحدى المطارات العالمية هو 0.15 خلال دقيقة واحدة. فأوجد:

1. احتمال هبوط طائرة في الدقيقة الثانية القادمة.
2. احتمال عدم هبوط أي طائرة خلال الثلاث دقائق القادمة.
3. أوجد التوقع والتباين لعدد الفترات الزمنية التي مدتها دقيقة واحدة اللازمة لهبوط أي طائرة في المطار.

تمرين (6.6): إذا كان احتمال أن يجتاز أي طالب الامتحان العملي في أي عام دراسي في معهد للتدريب المهني هو 0.60. وكان بإمكان الطلبة إعادة الامتحان حتى النجاح فيه، فأوجد؛

1. احتمال أن يجتاز طالب ما ذلك الامتحان في المحاولة الثانية خلال ثلاث سنوات.
2. احتمال أن يجتاز طالب ما ذلك الامتحان في المرة الثانية في أقل من أربع سنوات.
3. التوقع والتباين لعدد السنوات التي يتم خلالها اجتياز الامتحان في المحاولة الثانية.

تمرين (7.6): مستشفى للأمراض النفسية به 10 مرضى في قسم الحالات الخطرة، و 15 مريض في قسم الحالات العادية. تم اختيار 3 مرضى من المستشفى بصورة عشوائية، أوجد:

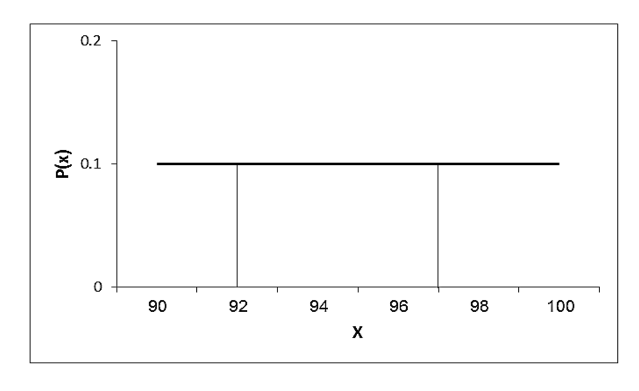
1. احتمال أن يكون كل المرضى الذين تم اختيارهم من الحالات الخطرة.
2. احتمال أن يكون هنالك مريضين من الحالات الخطرة في العينة.
3. احتمال أن يوجد مريضين خطرين على الأقل في العينة.
4. التوقع والتباين لعدد المرضى الخطرين.

تمرين (8.6): إذا كان معدل حجم التحميل لإحدى شركات الانترنت هو 6.5 جيجا بايت، وذلك خلال الخمسة دقائق الأخيرة من يوم الجمعة. فأوجد:

1. التوزيع الاحتمالي لحجم التحميل في الفترة المذكورة في يوم الجمعة.
2. احتمال أن يكون حجم التحميل في الفترة المذكورة هو 2 جيجا بايت.
3. احتمال أن يكون حجم التحميل هو 3 جيجا بايت على الأقل في الفترة المذكورة.
4. احتمال أن يكون حجم التحميل هو 2 جيجا بايت في العشرة دقائق الأخيرة من يوم الجمعة.
5. التوقع والتباين لحجم التحميل خلال الدقيقتين ونصف الأخيرة من يوم الجمعة.

تمرين (9.6): إذا كان احتمال أن يستغرق تحضير طبق رئيسي لمأدبة عشاء من 92 إلى 97 دقيقة ممثلا بالجزء المحدد كمستطيل في الشكل المرفق، فأوجد:

1. $P(92 < X < 97)$ مستخدما الشكل، وما هو التعليق المناسب؟
2. التوقع والتباين للزمن المستغرق لتحضير أي طبق رئيسي.



تمرين (10.6): إذا كان $Z \sim N(0, 1)$ ، فهل يمكن كتابة

$$P(-1.65 < Z < 1.65) = P(-1.65 < Z < 0) + P(0 < Z < 1.65)$$

تمرين (11.6): إذا كان الزمن من إجراء تحليل للسكر لآخر لغير مرضى السكر تحت سن 45 سنة يتبع التوزيع الطبيعي بمتوسط 30 شهر وانحراف معياري 6 أشهر. فأوجد احتمال أن يجري أحد الأشخاص تحليل السكر التالي:

1. خلال سنتين.   2. من سنتين لأربع سنوات.

تمرين (12.6): إذا علمت أن $Z \sim N(0, 1)$ ، وكان $P(-k < Z < k) = 0.95$ فأوجد قيمة k.

تمرين (13.6): إذا علمت أن 75% من الأشخاص الذين يملكون سيارات رياضية سريعة يحصلون على مخالفات لتجاوز السرعة المقررة شهريا في إحدى المدن. أوجد احتمال أن يحصل 420 شخص على الأقل ممن يملكون سيارات رياضية على مخالفات سرعة من ضمن عينة مكونة من 600 شخص يملكون سيارات رياضية. (استخدم تقريب توزيع ذي الحدين للتوزيع الطبيعي).

تمرين (14.6): إذا كان X متغير عشوائي يمثل نسبة مواد الإغاثة التي تستطيع فرق الأمم المتحدة إيصالها إلى إحدى المناطق المنكوبة في أفريقيا سنويا. وكان $X \sim Beta\ (x; 5, 3)$. فأوجد:

1. احتمال أن يتم إيصال 60% على الأقل من مواد الإغاثة خلال إحدى السنوات.
2. التوقع والتباين للنسبة مواد الإغاثة.

تمرين (15.6): إذا كان وصول الزبائن للخزينة في أحد الأسواق التجارية الكبيرة يعتبر أنه عملية بواسون بمتوسط 15 زبون في الساعة الواحدة. وكان X متغير عشوائي يمثل زمن وصول الزبائن للخزينة في الساعة، فأوجد:

1. احتمال أن لا يتواجد أي زبون في الخزينة خلال نصف ساعة.
2. التوقع والتباين للمتغير X .

الفصل السابع

توزيعات المعاينة
(Sampling Distributions)

1.7 مقدمة (Introduction)

2.7 توزيع المعاينة للوسط الحسابي للعينة (Sampling Distribution of the Sample Mean)

1.2.7 توزيع المعاينة لوسط واحد (Sampling Distribution of one Mean)

2.2.7 توزيع المعاينة للفرق بين وسطين
(Sampling Distribution of the difference between two Means)

3.7 توزيع المعاينة لنسب العينات (Sampling Distribution of the Sample Proportion)

1.3.7 توزيع المعاينة لنسبة واحدة (Sampling Distribution of one Proportion)

2.3.7 توزيع المعاينة للفرق بين نسبتين
(Sampling Distribution of the difference between two Proportions)

4.7 توزيع استيودنت t (Student's t Distribution)

5.7 توزيع مربع كاي χ^2 (Chi-Square Distribution)

6.7 توزيع فيشر F (Fisher's F Distribution)

7.7 تمارين الفصل السابع

1.7 **مقدمة** (Introduction)

في الفصل الأول، تم مناقشة مصادر جمع البيانات، وتم بشكل مختصر تعريف المجتمع والعينة وأنواع العينات وطرق اختيارها المختلفة. وفي هذا الفصل، سيتم توظيف ما تمت دراسته في علم الاحتمالات والتوزيعات الاحتمالية في عملية اختيار أو سحب العينات من المجتمع، وكذلك عرض التوزيعات الاحتمالية الخاصة ببعض المقاييس الإحصائية الهامة ومناقشة خواصها، ثم سيتم تقديم بعض التوزيعات الاحتمالية المتصلة التي ترتبط بتوزيع العينات.

إن تعاملنا مع الدراسات والأبحاث والتجارب التي تخص ظواهر الحياة المختلفة عادة ما يكون صعبا، إن لم يكن مستحيلا أحيانا، عندما تكون بيانات المجتمع ضخمة جدا أو منتشرة بشكل كبير أو أنها تتغير بصورة سريعة مع الزمن بحيث يتعذر التعامل معها بصورة مباشرة. لذلك فإن التعامل مع العينات هو الحل الأمثل للباحث أو للإحصائي بصورة عملية.

وسحب العينات يرتبط بالاحتمالات بصورة مباشرة لأن السحب عندما يكون بالإرجاع أو بدون إرجاع ويتم بصورة عشوائية فإن ذلك يؤكد ارتباط تلك العينات بقيم أو توزيعات احتمالية محددة. وعمليا، فإنه إذا كان المتغير العشوائي X مثلا يتبع توزيعا طبيعيا فإننا نقول أن المجتمع المعرّف عليه X هو مجتمع طبيعي، وإذا كان X يتبع توزيع ذي الحدين فإننا نقول أن المجتمع المعرّف عليه X يتبع توزيع ذي الحدين وهكذا. ومعرفتنا بتوزيع المجتمع يستدعي عادة معرفتنا بقيم معالم التوزيع الاحتمالي الخاص به، فعندما يكون $X \sim N(\mu, \sigma^2)$ مثلا فإن هذا يستوجب معرفة[1] قيم μ و σ^2 .

وهنا يجب توضيح أن تعاملنا مع المقاييس الإحصائية المختلفة سواء كانت مقاييس نزعة مركزية أو تشتت أو غيرها، سيكون من خلال توضيح ما تمثله؛ إما المجتمع وعندها يطلق على المقياس مصطلح **معلمة**، وإما العينة وعندها يطلق على المقياس مصطلح **إحصاءة**.

تعريف (1.7): **المعلمة** (Parameter): تعرّف أي قيمة عددية تصف مقياس أو خاصية للمجتمع بأنها معلمة.

تعريف (2.7): **الإحصاءة** (Statistic): تعرّف أي قيمة عددية تصف مقياس أو خاصية للعينة بأنها إحصاءة.

فمثلا؛ عندما يتم حساب مقياس الوسط الحسابي أو النسبة لمجموعة من البيانات فإن هذا الوسط أو النسبة تسمى معلمة إذا كانت البيانات تمثل المجتمع بأسره، ويسمى نفس المقياس إحصاءة إذا كانت البيانات تمثل عينة مسحوبة من ذلك المجتمع.

وتستخدم عادة رموز مختلفة لتوضيح الفرق بين المعلمة والإحصاءة، فمثلا؛ الرمز μ يُعبر عن الوسط الحسابي للمجتمع بينما يعبر $\bar{X}$ عن الوسط الحسابي للعينة، وكذلك الرمز σ^2 فإنه يعبر عادة عن تباين المجتمع بينما S^2 يعبر عن تباين العينة. وعموما فإن الكثير من المراجع العلمية تستخدم الرمز θ للدلالة بصورة عامة على مقياس المجتمع والرمز $\hat{\theta}$ للتعبير عن مقياس العينة "المُقدَر".

---

[1] في حال عدم معرفتنا بقيم معالم التوزيع الاحتمالي نلجأ لتقديرها من خلال العينة، وهذا يندرج تحت مفهوم نظرية التقدير.

إضافة إلى ذلك، فإن مصطلح إحصاءة يستخدم أيضا للتعبير عن المتغير العشوائي نفسه، فمثلا يقال (كما سنرى لاحقا) أن $\bar{X}$ هو متغير عشوائي له دالة توزيع احتمالي $P_{\bar{X}}(x)$ وهو، أي $\bar{X}$ ، يمثل أيضا إحصاءة للعينة (أو العينات) التي تم حساب قيمته منها، ويقال أن $\bar{X}$ في هذه الحالة يكون له **توزيع معاينة**.

ويبرز مفهوم ظهور توزيع احتمالي لمقياس خاص بالعينة من واقع أنه يمكن سحب عدة عينات من مجتمع واحد[1] بحيث تختلف هذه العينات عن بعضها البعض (من حيث ما تحويه من قيم البيانات)، وبالتالي حساب أي مقياس باستخدام كل عينة سينشأ عنه قيمة مختلفة. وهكذا يتكون لدينا مجموعة مختلفة من قيم هذا المقياس والتي بدورها تمثل متغير عشوائي (ولنرمز له بالرمز X^*)، وهذا المتغير سيرتبط بدالة احتمالية لأن سحب العينات من المجتمع تم بصورة عشوائية، وبالتالي نقول أن المتغير العشوائي X^* له توزيع معاينة يعتمد على طبيعة المجتمع وطريقة سحب العينات وحجمها وغير ذلك من الأمور. وقد يتبع توزيع X^* نفس توزيع المجتمع الأصلي أو يتبع توزيع آخر اعتمادا على طبيعة التوزيع الأصلي ونوع المقياس المحسوب من العينات.

ولتوضيح الصورة أكثر، نأخذ الوسط الحسابي كمثال فيكون؛ $X^* = \bar{X}$ وهذا يعني أنه سيتم حساب الوسط الحسابي لكل عينة مسحوبة من المجتمع، فينتج لدينا عدة متوسطات. هذه المتوسطات تمثل بحد ذاتها قيم لمتغير عشوائي (لأن كل عينة سحبت عشوائيا من المجتمع) يكون له توزيع احتمالي يعرف بتوزيع المعاينة.

تعريف (3.7): **توزيع المعاينة** (Sampling Distribution): توزيع المعاينة هو التوزيع الاحتمالي للإحصاءة، منفصلا كان أو متصلا.

ملاحظة: يطلق على الانحراف المعياري الخاص بأي توزيع معاينة مصطلح **الخطأ المعياري** (Standard Error, (*S.E*)) للإحصاءة، وذلك لأن قيمته تمثل "تقديرا" للقيمة الحقيقية لتشتت القيم عن وسطها الحسابي، وهذا التقدير لابد أن يعتريه ابتعاد عن القيمة الحقيقة (خطأ تقدير).

وبصورة عامة، فإن دراسة توزيعات المعاينة يندرج تحت مسمى دراسة **نظرية المعاينة** (Sampling Theory)، والتي تهتم بالتعامل مع المجتمعات والعينات المسحوبة منها. وتمثل هذه النظرية الأساس الذي يقوم عليه علم الاستدلال الإحصائي.

ومن المفاهيم التي قد تختلط على البعض، هي اعتبار مجموعة من المتغيرات العشوائية عينة مسحوبة من المجتمع، ولتوضيح هذه النقطة نقدم التعريف التالي.

تعريف (4.7): **المتغيرات العشوائية كعينة من المجتمع:** يقال أن المتغيرات العشوائية $X_1, X_2, \ldots, X_n$ هي عينة عشوائية حجمها n إذا كانت:

1. متغيرات عشوائية مستقلة.
2. كل X_i له نفس التوزيع الاحتمالي، $i = 1, 2, \ldots, n$.

فمثلا، إذا ما اعتبرنا أن المتغير العشوائي X يمثل معدلات (درجات) الطلبة في إحدى الجامعات العربية، والتي يفترض أن تتبع التوزيع الطبيعي بمتوسط μ وتباين σ^2 ، وكان X_1 يمثل معدل طلبة كلية الآداب، و X_2 يمثل

[1] أو حتى من عدة مجتمعات.

معدل طلبة كلية العلوم، X_3 يمثل معدل طلبة كلية الهندسة، ...، وهكذا إلى باقي الكليات. فإن هذه المتغيرات تكون كلها مستقلة عن بعضها البعض (بحكم استقلال الكليات)، ويمكن على سبيل المثال أن نسحب عينة مكونة من $n = 5$ كلية من هذه الجامعة بحيث تكون $X_1, X_2, \dots, X_5$ هي متغيرات مستقلة عن بعضها وكل منها له توزيع طبيعي بمتوسط μ وتباين σ^2 . وتكون $x_1, x_2, \dots, x_5$ هي قيم تلك المتغيرات على الترتيب.

وبصورة عملية، فإن طريقة سحب العينة من المجتمع هي التي تُضفي على حجم المجتمع سمته الرئيسية، بمعنى أنه إذا تم سحب العينات بالإرجاع (وهذا يجعل لكل مفردة فرصة الظهور أكثر من مرة)، فإنه يمكن اعتبار المجتمع في هذه الحالة غير منتهي أو كبير جدا، وأما إذا تم سحب العينات بدون إرجاع فإن المجتمع عندئذ يعتبر محدودا.

وسنقوم في البندين القادمين (2.7) و (3.7) بعرض توزيع المعاينة (وعلاقته بالتوزيع الطبيعي) لبعض المقاييس المتداولة مثل الوسط الحسابي والنسبة. ثم نقوم بتعريف بعض توزيعات المعاينة الهامة والمرتبطة أيضا، كما سنرى تباعا، بالتوزيع الطبيعي وهي توزيع t ، توزيع مربع كاي، وتوزيع F .

2.7 توزيع المعاينة للوسط الحسابي للعينة (Sampling Distribution of the Sample Mean)

إن الوسط الحسابي كمقياس للنزعة المركزية يعتبر من أبسط وأهم المقاييس المستخدمة في كل مجالات الحياة نظرا لسهولة استخدامه وتفسير ما يمثله أو يصفه من البيانات. وفيما يلي سنستعرض الآلية التي يتم من خلالها سحب العينات وحساب المتوسطات لها، بغية تكوين توزيع المعاينة لهذه الأوساط.

لنفرض أنه لدينا مجتمع حجمه N ومفرداته هي $x_1, x_2, \dots, x_N$ ، وله التوزيع الاحتمالي $f_X(x)$ (معلوما كان أو مجهولا)، بمتوسط يساوي μ وتباين يساوي σ^2 ، وتم سحب[1] كل العينات الممكنة التي حجمها n من هذا المجتمع (بالإرجاع أو بدون إرجاع). ولنفرض أيضا أننا قمنا بحساب الوسط الحسابي لكل عينة جزئية فكانت قيم هذه الأوساط هي $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots$ ، حيث يكون عدد هذه الأوساط هو عدد كل العينات الممكن سحبها من المجتمع. ما نريد الوصول إليه الآن هو معرفة التوزيع الاحتمالي للمتغير الجديد $\bar{X}$ والذي يأخذ القيم $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots$ والتي سيكون لها المتوسط $\mu_{\bar{X}}$ والتباين $\sigma_{\bar{X}}^2$ ، ونرغب أيضا بمعرفة علاقة الوسط والتباين للمتغير $\bar{X}$ بالوسط والتباين للمتغير الأصلي X . والبند القادم (1.2.7) يعرض لنا النظرية الأساسية التي تنظم هذه العلاقات و توّضح توزيع $\bar{X}$.

1.2.7 توزيع المعاينة لوسط واحد (Sampling Distribution of one Mean)

لنقوم أولا بتناول المثال التالي الذي يوضح مفهوم عملية المعاينة وحساب المتوسطات للعينات.

مثال (1.7): لنفرض أنه لدينا المتغير العشوائي X الذي له دالة الكتلة الاحتمالية:

$$f_X(x) = \frac{1}{5} \quad , \quad x = 1, 2, \dots, 5$$

---

[1] السحب بأسلوب يتناسب مع طبيعة المجتمع والدراسة المراد القيام بها.

وأن هذا التوزيع هو خاص بأحد المجتمعات. في هذه الحالة نستطيع القول بأن متوسط هذا المجتمع هو

$$\mu = E(X) = \sum_{x=1}^{5} x\, f(x) = 1 \times \tfrac{1}{5} + 2 \times \tfrac{1}{5} + \cdots + 5 \times \tfrac{1}{5} = 3$$

وتباين هذا المجتمع هو

$$\sigma^2 = Var(X) = \sum_{x=1}^{5} (x-\mu)^2 f(x) = E(X^2) - (E(X))^2 = 11 - 3^2 = 2$$

حيث

$$E(X^2) = \sum_{x=1}^{5} x^2 f(x) = 1^2 \times \tfrac{1}{5} + 2^2 \times \tfrac{1}{5} + \cdots + 5^2 \times \tfrac{1}{5} = 11$$

الآن لنفرض أننا نريد سحب كل العينات الممكنة ذات الحجم $n = 2$ ، مثلا، بالإرجاع من هذا المجتمع، ومن ثم حساب الوسط الحسابي لكل عينة مسحوبة. عندئذ سنحصل على النتائج المبينة في جدول (1.7)، حيث أن العدد الكلي للعينات ذات الحجم $n = 2$ والممكن سحبها من المجتمع الذي حجمه $N = 5$ هو $N^n = 5^2 = 25$ عينة.

جدول (1.7): الأوساط الحسابية للعينات المسحوبة في المثال (1.7).

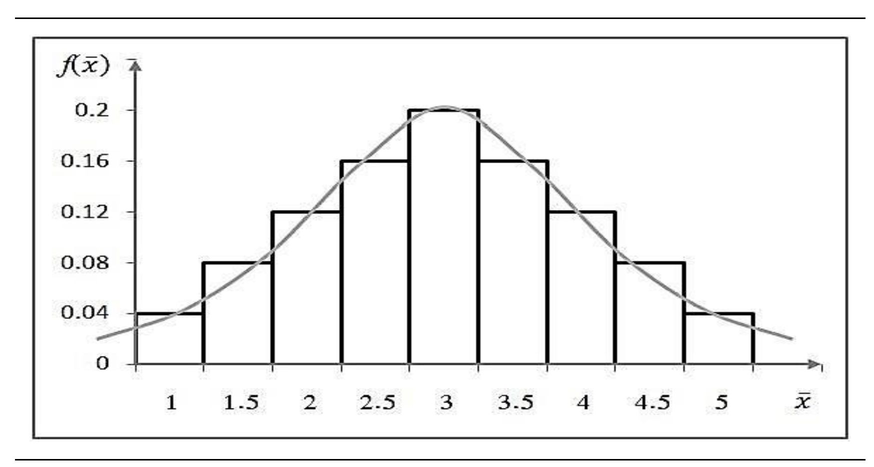
ترتيب العينة	العينة	الوسط الحسابي للعينة
1	1،1	1
2	1،2	1.5
3	1،3	2
4	1،4	2.5
5	1،5	3
6	2،1	1.5
7	2،2	2
8	2،3	2.5
9	2،4	3
10	2،5	3.5
11	3،1	2
12	3،2	2.5
13	3،3	3
14	3،4	3.5
15	3،5	4
16	4،1	2.5
17	4،2	3
18	4،3	3.5
19	4،4	4
20	4،5	4.5
21	5،1	3
22	5،2	3.5
23	5،3	4
24	5،4	4.5
25	5،5	5

ويمكن تلخيص الجدول (1.7) إذا تم رصد قيم الوسط الحسابي للعينة وتكراراتها المناظرة، كما هو مبين في جدول (2.7). ولاحظ أن العامود الأخير في جدول (2.7) يحتوي على الاحتمالات المناظرة لقيم الوسط الحسابي للعينة، وبالتالي يكون الجدول ممثلا للتوزيع الاحتمالي للوسط الحسابي للعينة والذي يعرف بتوزيع المعاينة للوسط الحسابي.

جدول (2.7): توزيع المعاينة للوسط الحسابي للمثال (1.7).

الوسط الحسابي للعينة $\bar{x}$	**التكرار f**	$f(\bar{x})$
1	1	1/25
1.5	2	2/25
2	3	3/25
2.5	4	4/25
3	5	5/25
3.5	4	4/25
4	3	3/25
4.5	2	2/25
5	1	1/25
المجموع	25	1

ويمكن رسم المدرج الاحتمالي (أو المنحنى الاحتمالي) لتوزيع المعاينة في جدول (2.7) كما هو موضح في الشكل (1.7)، حيث يمكن ملاحظة أنه يأخذ شكلا مطابق للتوزيع الطبيعي، رغم أن توزيع المفردات الأصلي (المجتمع) هو توزيع منتظم.



شكل (1.7): المدرج والمنحنى الاحتمالي لتوزيع المعاينة للوسط الحسابي للمثال (1.7).

الآن سنقوم بحساب الوسط الحسابي والتباين لتوزيع المعاينة للوسط $\bar{X}$ بالصورة التالية:

$$\mu_{\bar{X}} = E(\bar{X}) = \sum_{x} \bar{x}\, f(\bar{x}) = (1) \times \frac{1}{25} + (1.5) \times \frac{2}{25} + \cdots + (5) \times \frac{1}{25} = 3 = \mu$$

وكذلك

$$E(\bar{X}^2) = \sum_{x} \bar{x}^2\, f(\bar{x}) = (1)^2 \times \frac{1}{25} + (1.5)^2 \times \frac{2}{25} + \cdots + (5)^2 \times \frac{1}{25} = 10$$

وبالتالي يكون

$$\sigma_{\bar{X}}^2 = Var(\bar{X}) = E(\bar{X}^2) - (E(\bar{X}))^2 = 10 - 3^2 = 1$$

ولاحظ أننا نستطيع كتابة تباين توزيع المعاينة للوسط $\bar{X}$ بالصورة

$$\sigma_{\bar{X}}^2 = 1 = \frac{2}{2} = \frac{\sigma^2}{n}$$

بمعنى أن الوسط الحسابي لتوزيع المعاينة للوسط $\bar{X}$ يساوي الوسط الحسابي للمجتمع الأصلي، والتباين لتوزيع المعاينة للوسط $\bar{X}$ يساوي تباين المجتمع الأصلي مقسوما على حجم العينة، وذلك عندما تكون المعاينة (السحب) بالإرجاع. وهذه النتيجة هي في الواقع نظرية عامة تنطبق على أي حالة يتم فيها سحب كل العينات الممكنة ذات الحجم n بالإرجاع من مجتمع حجمه N وله توزيع احتمالي $f(x)$ ، وهذا ما تنص عليه النظرية التالية.

نظرية (1.7): إذا تم سحب كل العينات العشوائية الممكنة ذات الحجم n بالإرجاع من مجتمع منتهي حجمه N، (حيث $n \leq N$)، وله الوسط الحسابي μ والتباين σ^2 ، عندها يكون توزيع المعاينة للوسط $\bar{X}$ موزعا بتوزيع طبيعي تقريبا بمتوسط $\mu_{\bar{X}} = \mu$ وتباين $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$ ، وبالتالي يكون المتغير العشوائي

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

هو متغير يتبع التوزيع الطبيعي المعياري.

وأما في الحالة التي يكون فيها سحب العينات بدون إرجاع فإن العلاقة بين تباين توزيع المعاينة للوسط $\bar{X}$ وتباين المجتمع تتغير مع بقاء $\mu_{\bar{X}} = \mu$ كما توضح النظرية التالية.

نظرية (2.7): إذا تم سحب كل العينات العشوائية الممكنة ذات الحجم n بدون إرجاع من مجتمع منتهي حجمه N ، (حيث $n \leq N$)، وله الوسط الحسابي μ والتباين σ^2 ، عندها يكون توزيع المعاينة للوسط $\bar{X}$ موزعا بتوزيع طبيعي تقريبا بمتوسط $\mu_{\bar{X}} = \mu$ وتباين[1] $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}\,\frac{N-n}{N-1}$.

مثال (2.7): من بيانات المجتمع التالي 7 ,6 ,5 ,3 ,3 ,2 أوجد الوسط الحسابي والتباين لتوزيع المعاينة للوسط $\bar{X}$ إذا ما تم سحب كل العينات الممكنة ذات الحجم $n = 3$ بدون إرجاع.

الحل:

يمكن اعتمادا على النظرية (2.7) حساب الوسط والتباين لتوزيع $\bar{X}$ من خلال علاقتهما بوسط وتباين المجتمع، إلا أننا سنقوم بحساب قيمهما بالطريقة التقليدية لإضفاء المزيد من التوضيح العملي.

إن عدد العينات العشوائية ذات الحجم $n = 3$ والممكن سحبها من $N = 6$ مفردة من مفردات المجتمع بدون إرجاع هو $C_n^N = C_3^6 = 20$ عينة . وهذه العينات هي موضحة في جدول (3.7) بالإضافة لأوساطها الحسابية.

---

[1] حيث يسمى الحد $\frac{N-n}{N-1}$ **بمعامل التصحيح للمجتمع المنتهي** (Finite Population Correction Factor).

جدول (3.7): الأوساط الحسابية للعينات المسحوبة بدون إرجاع في المثال (2.7).

ترتيب العينة	العينة	الوسط الحسابي للعينة	ترتيب العينة	العينة	الوسط الحسابي للعينة
1	2,3,3	2.67	11	3,3,5	3.67
2	2,3,5	3.33	12	3,3,6	4
3	2,3,6	3.67	13	3,3,7	4.33
4	2,3,7	4	14	3,5,6	4.67
5	2,3,5	3.33	15	3,5,7	5
6	2,3,6	3.67	16	3,6,7	5.33
7	2,3,7	4	17	3,5,6	4.67
8	2,5,6	4.33	18	3,5,7	5
9	2,5,7	4.67	19	3,6,7	5.33
10	2,6,7	5	20	5,6,7	6

والجدول (4.7) يوضح توزيع المعاينة للوسط الحسابي $\bar{X}$.

جدول (4.7): توزيع المعاينة للوسط الحسابي للمثال (2.7).

الوسط الحسابي للعينة $\bar{x}$	التكرار f	$f(\bar{x})$
2.67	1	0.05
3.33	2	0.10
3.67	3	0.15
4	3	0.15
4.33	2	0.10
4.67	3	0.15
5	2	0.10
5.33	3	0.15
6	1	0.05
المجموع	20	1

ويمكن ملاحظة أن

$$\mu_{\bar{X}} = \sum_{x} \bar{x}\, f(\bar{x}) = (2.67) \times (0.05) + (3.33) \times (0.10) + \cdots + (6) \times (0.05) = 4.35$$
$$\cong 4.33 = \mu$$

وكذلك

$$\sigma_{\bar{X}}^2 = E(\bar{X}^2) - (E(\bar{X}))^2 = 19.59 - (4.35)^2 = 0.67 \cong \left(\frac{3.22}{3}\,\frac{6-3}{6-1} = 0.64\right)$$

وهذا الاختلاف الطفيف في قيم الوسط الحسابي والتباين بين توزيع المعاينة والمجتمع هو نتيجة خطأ التدوير إلى خانتين، وإذا ما تم استخدام القيم العشرية كاملة فإن قيم الوسط والتباين لكل من توزيع المعاينة والمجتمع ستكون متطابقة تماما حسب النظرية.

مثال (3.7): إذا علمت أن أطوال الطلبة في إحدى الكليات تتوزع بتوزيع طبيعي بمتوسط 1.73 متر وانحراف معياري 0.076 متر، وتم سحب كل العينات الممكنة ذات الحجم $n = 25$ طالب من المجتمع الذي حجمه $N =$ 3000 طالب فأوجد:

1. الوسط الحسابي والانحراف المعياري لتوزيع المعاينة للأوساط الناتجة عن سحب العينات العشوائية بالإرجاع وبدون إرجاع.
2. احتمال سحب عينات لها أوساط بين 1.70 متر و 1.73 متر إذا كان السحب بالإرجاع.

الحل:

1. من النظريتين (1.7) و (2.7) لدينا $\mu_{\bar{X}} = \mu = 1.73$ متر

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.076}{\sqrt{25}} = 0.015 \text{ متر}$$

وذلك عندما يكون السحب بالإرجاع. أما عندما يكون السحب بدون إرجاع فإن الانحراف المعياري لتوزيع المعاينة يساوي

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{0.076}{\sqrt{25}}\sqrt{\frac{3000-25}{3000-1}} \cong 0.015 \text{ متر}$$

وهذا يعني أن توزيع المعاينة للوسط $\bar{X}$ يتوزع بتوزيع طبيعي تقريبا بمتوسط 1.73 متر وتباين 0.0002 متر.

2. من النظرية (1.7) يمكن تعريف المتغير العشوائي Z بالصورة

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-1.73}{0.015}$$

وبالتالي فإن الاحتمال المطلوب هو

$$P(1.70 < \bar{X} < 1.73) = P\left(\frac{1.70-1.73}{0.015} < \frac{\bar{X}-1.73}{0.015} < \frac{1.73-1.73}{0.015}\right)$$
$$= P(-2 < Z < 0) = P(Z < 0) - P(Z < -2) = 0.5 - 0.0228$$
$$= 0.4772$$

لاحظنا في المثال (1.7) كيف أن توزيع المعاينة للوسط الحسابي يتبع توزيعا طبيعيا رغم أن التوزيع الأصلي للمجتمع كان يتبع توزيعا منتظما. هذه النتيجة أو الظاهرة تم تعميمها في الواقع لتشمل أي توزيع للمجتمع. بمعنى أنه إذا تم سحب عينات عشوائية من المجتمع فإن توزيع المعاينة للوسط الحسابي لهذه العينات (والتي يكون عددها كبير بدرجة كافية) سيتبع توزيع طبيعي بغض النظر عن نوع التوزيع الأصلي للمجتمع، وهذا ما تنص

عليه نظرية النهاية المركزية التي نعرضها فيما يلي بصيغتين تشتركان في المفهوم وتختلفان في طريقة عرض المعطيات.

نظرية (3.7): **نظرية النهاية المركزية** (Central Limit Theorem)

الصيغة (أ):

إذا تم سحب عينات حجم كل منها n من مجتمع كبير أو غير محدود له متوسط μ وتباين σ^2، فإن توزيع المعاينة للأوساط الحسابية لهذه العينات، ممثلة بالمتغير العشوائي $\bar{X}$ ، يتبع توزيع طبيعي تقريبا بمتوسط $\mu_{\bar{X}} = \mu$ وتباين $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$ ، ويكون

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

هو متغير عشوائي يتبع التوزيع الطبيعي المعياري.

الصيغة (ب):

إذا كانت $X_1, X_2, \dots, X_n$ هي متغيرات عشوائية مستقلة ولها نفس التوزيع الاحتمالي بمتوسط $E(X_i) = \mu$ وتباين $Var(X_i) = \sigma^2$ ، $i = 1, 2, \dots, n$ ، وكان

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad , \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

فإن توزيع Z يقترب من التوزيع الطبيعي المعياري كلما زادت قيمة n ، (أي $n \to \infty$).

ويمكن عمليا الاسترشاد بالنقاط التالية فيما يخص عدد العينات المسحوبة أو حجم العينة:

1. إذا كان $n < 15$ يعتبر توزيع المعاينة للوسط $\bar{X}$ طبيعيا فقط إذا كان توزيع المجتمع طبيعيا.
2. إذا كان $15 < n \le 40$ يعتبر توزيع المعاينة للوسط $\bar{X}$ طبيعيا تقريبا إلا إذا كان توزيع البيانات في المجتمع يحوي قيما متطرفة أو كان به التواء حاد.
3. إذا كان $40 < n \le 100$ يعتبر توزيع المعاينة للوسط $\bar{X}$ طبيعيا تقريبا إلا إذا كان توزيع البيانات في المجتمع يحوي قيما متطرفة أو كان لها أكثر من منوال.
4. إذا كان $n > 100$ فإن توزيع المعاينة للوسط $\bar{X}$ سيكون دائما طبيعيا.

2.2.7 توزيع المعاينة للفرق بين وسطين

(Sampling Distribution of the difference between two Means)

إن مفهوم التعامل مع توزيع المعاينة للوسط الحسابي للعينة يمكن التوسع فيه ليشمل التعامل مع مجتمعين، بحيث يتم سحب عينات من كل مجتمع على حده. ويكون المتغير العشوائي المراد التعامل معه وإيجاد متوسط وتباين توزيعه هو الفرق بين متوسطات العينات كما يتضح من النظرية التالية.

نظرية (4.7): إذا تم سحب عينات عشوائية مستقلة حجمها n_1 و n_2 من مجتمعين كبيرين لهما المتوسطات μ_1 و μ_2 والتباينات σ_1^2 و σ_2^2 على الترتيب، فإن توزيع المعاينة للفرق بين الوسطين $\bar{X}_1$ و $\bar{X}_2$ ، (والتي تعتبر متغيرات عشوائية تمثل أوساط العينات المسحوبة من المجتمع الأول والثاني على الترتيب)، يكون طبيعيا تقريبا بمتوسط وتباين معطى بالصورة

$$E(\bar{X}_1 - \bar{X}_2) = \mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$$

$$Var(\bar{X}_1 - \bar{X}_2) = \sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

ويكون

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}}$$

هو متغير عشوائي يتبع التوزيع الطبيعي المعياري.

مثال (4.7): لنفرض أنه لدينا المجتمعين[1] I و II اللذان يحتويان البيانات التالية:

المجتمع I: 3 ، 5 ، 7 ، و المجتمع II: 0 ، 3

ونريد سحب كل العينات الممكنة بالإرجاع من المجتمع الأول ذات الحجم $n_1 = 2$ ، وذات الحجم $n_2 = 3$ من المجتمع الثاني.

الحل:

من المجتمع I :

$$\mu_1 = \frac{3 + 5 + 7}{3} = 5$$

$$\sigma_1^2 = \frac{(3-5)^2 + (5-5)^2 + (7-5)^2}{3} = \frac{8}{3}$$

ومن المجتمع II :

$$\mu_2 = \frac{0 + 3}{2} = 1.5$$

$$\sigma_2^2 = \frac{(0-1.5)^2 + (3-1.5)^2}{2} = \frac{9}{4}$$

ويكون عدد العينات في المجتمع الأول هو $N_1^{n_1} = 3^2 = 9$ عينات، و عدد العينات في المجتمع الثاني هو $N_2^{n_2} = 2^3 = 8$ عينات. وهكذا تكون متوسطات العينات كما هو موضح في جدول (5.7).

جدول (5.7): الأوساط الحسابية للعينات المسحوبة في المثال (4.7).

[1] لاحظ من النظرية (4.7) أن المجتمعين لابد أن تكون أحجامهما كبيرة، إلا أننا نستخدم هذه المجتمعات الصغيرة الحجم في هذا المثال لتوضيح طريقة التعامل الرياضي بصورة مبسطة.

من المجتمع I			من المجتمع II		
ترتيب العينة	العينة	الوسط الحسابي للعينة	ترتيب العينة	العينة	الوسط الحسابي للعينة
1	3,3	3	1	0,0,0	0
2	3,5	4	2	0,0,3	1
3	3,7	5	3	0,3,0	1
4	5,3	4	4	3,0,0	1
5	5,5	5	5	0,3,3	2
6	5,7	6	6	3,0,3	2
7	7,3	5	7	3,3,0	2
8	7,5	6	8	3,3,3	3
9	7,7	7			

والآن نقوم بحساب الفروق بين قيم هذه المتوسطات $(\bar{X}_1 - \bar{X}_2)$ ، والجدول (6.7) يوضح هذه الفروق.

جدول (6.7): الفروق بين المتوسطات المحسوبة لعينات المجتمعين في المثال (4.7).

	$\bar{X}_1$								
$\bar{X}_2$	3	4	5	4	5	6	5	6	7
0	3	4	5	4	5	6	5	6	7
1	2	3	4	3	4	5	4	5	6
1	2	3	4	3	4	5	4	5	6
1	2	3	4	3	4	5	4	5	6
2	1	2	3	2	3	4	3	4	5
2	1	2	3	2	3	4	3	4	5
2	1	2	3	2	3	4	3	4	5
3	0	1	2	1	2	3	2	3	4

وبالتالي يمكن تلخيص قيم هذه الفروق في الجدول السابق وتكوين توزيع المعاينة للفرق بين الأوساط $(\bar{X}_1 - \bar{X}_2)$ كما هو الحال في جدول (7.7).

جدول (7.7): توزيع المعاينة للفرق بين أوساط العينات للمثال (4.7).

$\bar{X}_1 - \bar{X}_2$	التكرار f	$f(\bar{X}_1 - \bar{X}_2)$
0	1	1/72
1	5	5/72
2	12	12/72
3	18	18/72
4	18	18/72
5	12	12/72
6	5	5/72
7	1	1/72
المجموع	72	1

وهكذا يمكننا حساب الوسط الحسابي والتباين لتوزيع المعاينة للفرق بين الأوساط بالصورة

$$\mu_{\bar{X}_1-\bar{X}_2} = \sum (\bar{X}_1 - \bar{X}_2)\, f(\bar{X}_1 - \bar{X}_2) = 0 \times \frac{1}{72} + \cdots + 7 \times \frac{1}{72} = \frac{7}{2}$$

ولاحظ أن

$$\mu_{\bar{X}_1-\bar{X}_2} = \frac{7}{2} = 5 - \frac{3}{2} = \mu_1 - \mu_2$$

وكذلك يكون

$$\sigma^2_{\bar{X}_1-\bar{X}_2} = \sum \left((\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)\right)^2 f(\bar{X}_1 - \bar{X}_2)$$
$$= \left(0 - \tfrac{7}{2}\right)^2 \times \tfrac{1}{72} + \cdots + \left(7 - \tfrac{7}{2}\right)^2 \times \tfrac{1}{72} = \frac{25}{12}$$

وأيضا هنا يمكننا ملاحظة أن

$$\sigma^2_{\bar{X}_1-\bar{X}_2} = \frac{25}{12} = \frac{\frac{8}{3}}{2} + \frac{\frac{9}{4}}{3} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

مثال (5.7): يوجد في أحد المتاجر نوعان من المصابيح الكهربية، النوع A والذي له متوسط عمر استهلاكي يقدر بـ 1400 ساعة وانحراف معياري 200 ساعة، والنوع B والذي له متوسط عمر استهلاكي يقدر بـ 1200 ساعة بانحراف معياري 100 ساعة. تم سحب عينتين عشوائيتين حجم كل منهما 125 مصباح من كل نوع وتم اختبارها. ما احتمال أن مصابيح النوع A سيكون لها متوسط عمر استهلاكي أكبر من مصابيح النوع B بـ 160 ساعة؟.

الحل:

لنفرض أن $\bar{X}_A$ و $\bar{X}_B$ تمثل متوسط العمر الاستهلاكي للعينات من نوع A و B على الترتيب. فيكون متوسط الفرق بين وسطي العينتين هو

$$\mu_{\bar{X}_A-\bar{X}_B} = \mu_A - \mu_B = 1400 - 1200 = 200 \text{ ساعة}$$

والانحراف المعياري للفرق بين وسطي العينتين هو

$$\sigma_{\bar{X}_A-\bar{X}_B} = \sqrt{\frac{\sigma_A^2}{n_A}+\frac{\sigma_B^2}{n_B}} = \sqrt{\frac{(200)^2}{125}+\frac{(100)^2}{125}} = 20 \text{ ساعة}$$

وطبقا للنظرية (4.7) يكون المتغير العشوائي

$$Z = \frac{(\bar{X}_A-\bar{X}_B)-(\mu_A-\mu_B)}{\sigma_{\bar{X}_A-\bar{X}_B}} = \frac{(\bar{X}_A-\bar{X}_B)-200}{20}$$

موزعا بتوزيع طبيعي معياري، وبالتالي يكون المطلوب هو حساب الاحتمال:

$$P(\bar{X}_A-\bar{X}_B>160) = P\left(\frac{(\bar{X}_A-\bar{X}_B)-(\mu_A-\mu_B)}{\sigma_{\bar{X}_A-\bar{X}_B}} > \frac{160-200}{20}\right) = P(Z>-2)$$

$$= 1-P(Z\le -2) = 1-0.0228 = 0.9772$$

ومن البديهي أن يكون هذا الاحتمال كبيرا جدا نظرا لأن مصابيح النوع A لها متوسط عمر استهلاكي أكبر من مصابيح النوع B .

3.7 توزيع المعاينة لنسب العينات (Sampling Distribution of the Sample Proportion)

في كثير من الدراسات والتجارب العشوائية قد نتعامل مع متغيرات وصفية لا تأخذ قيما رقمية حقيقية بصورة مباشرة. في هذه الحالة لا معنى لحساب الوسط الحسابي[1] لتلك المتغيرات الوصفية ولا حتى التباين. وبالتالي فإن توزيع المعاينة لمقياس مثل الوسط الحسابي لتلك المتغيرات سيكون من غير المنطقي التعامل معه بصورة مباشرة.

لذلك فإنه يمكن الاعتماد على مقياس آخر مثل النسبة للتعامل بصورة أكثر منطقية من هذا النوع من المتغيرات. فمثلا إذا كان المتغير العشوائي يمثل النوع أو فئة الدم أو اللون، فإن التوزيع الاحتمالي لهذا المتغير سيعتمد على النسبة (أو التكرار النسبي) للمفردات التي تتمتع بالصفة المطلوبة (صفة المتغير العشوائي). ولمزيد من التوضيح نأخذ المثال التوضيحي التالي:

لنفرض أنه تم رمي عملة معدنية n من المرات، وكان X متغير عشوائي يمثل عدد مرات ظهور الصورة في الـ n رمية. من الواضح أن المتغير X سيتبع توزيع ذي الحدين، إلا أننا لسنا مهتمين هنا بالمتغير X بل بالإحصاءة التي تمثل نسبة مفردات العينة (التي حجمها n) والتي تمثل حالة النجاح وهي ظهور الصورة.

---

[1] حيث أن المتغيرات الوصفية، كما أشرنا سابقا، يتم التعبير عنها برموز رقمية لا معنى منطقي للتعامل معها باستخدام الكثير من المقاييس الإحصائية، فمثلا متغير النوع الذي يأخذ القيم 1، 2، 2 حيث يرمز 1 للذكر و 2 للأنثى يكون متوسطه الحسابي هو 1.67 وهي قيمة لا تعبر أو تلخص أي صفة في المتغير.

هذه الإحصاءة، والتي سيرمز لها بالرمز p ، تمثل نسبة العينة، وقيمتها هي $p = \frac{x}{n}$. ويُرمز لمعلمة المجتمع (نسبة المجتمع) بالرمز P . وسنقوم فيما يلي بمناقشة كيفية التعامل مع توزيع المعاينة لنسبة العينة وحساب الوسط والتباين عند التعامل مع مجتمع واحد ومجتمعين.

1.3.7 توزيع المعاينة لنسبة واحدة (Sampling Distribution of one Proportion)

لنفرض أنه لدينا مجتمع كبير أو غير محدود، ويتوزع بتوزيع ذي الحدين بالمعالم N وP، ولنفرض أنه يمكن سحب كل العينات الممكنة ذات الحجم n من هذا المجتمع، ومن ثمة اعتبار p هي نسبة النجاح في كل العينة. عندئذ يمكن الحصول على توزيع المعاينة للنسبة p والتي سيكون الوسط الحسابي والتباين له معرف بالصورة:

$$\mu_p = p \quad , \quad \sigma_p^2 = \frac{p(1-p)}{n}$$

وعندما يكون حجم العينات (أو عدد المحاولات) $n \geq 30$ فإن توزيع المعاينة للنسبة سيقترب من التوزيع الطبيعي.

مثال (6.7): إذا كانت نسبة المعيب في إنتاج إحدى آلات تصنيع المراوح الكهربائية هو 2%. فما هو احتمال أن يوجد في عينة مكونة من 400 مروحة 3% أو أكثر وحدة معيبة؟.

الحل:

لنفرض أن نسبة المراوح المعيبة في العينة هو p ، عندها يمكن حساب الوسط و الانحراف المعياري لهذه النسبة كالتالي

$$\mu_p = p = 0.02 \quad , \quad \sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.02 \times 0.98}{400}} = 0.007$$

أي أن النسبة p تتبع توزيعا طبيعيا بمتوسط 0.02 و تباين 0.00005 ، والمطلوب هو حساب الاحتمال

$$P(P \geq 0.03) = P\left(\frac{P-\mu_p}{\sigma_p} \geq \frac{0.03-0.02}{0.007}\right) = P(Z \geq 1.43) = 1 - P(Z < 1.43)$$
$$= 1 - 0.9236 = 0.076$$

في بعض الحالات قد نتعامل مع مجتمعات حجمها محدود أو صغير، ويتم في نفس الوقت سحب العينات من المجتمع بدون إرجاع، عندئذ يتم حساب تباين الإحصاءة P بالصورة:

$$\sigma_p^2 = \frac{p(1-p)}{n} \frac{(N-n)}{(N-1)}$$

أي يتم في هذه الحالة استخدام معامل التصحيح للمجتمع المنتهي والذي سبقت الإشارة إليه.

2.3.7 توزيع المعاينة للفرق بين نسبتين
(Sampling Distribution of the difference between two Proportions)

كما كان الحال مع الأوساط الحسابية للعينات، فإنه يمكن إيجاد توزيع المعاينة للفرق بين نسبتي عينتين مسحوبتين من مجتمعين مستقلين. فإذا كان هنالك مجتمعين يتبعان توزيع ذي الحدين بالمعالم N_1 و P_1 و N_2 و P_2 على الترتيب، فإن الوسط الحسابي والتباين للفرق بين نسبتي عينتين (حجمهما n_1 و n_2) مسحوبتين من المجتمعين الأول والثاني يعطى بالصورة:

$$\mu_{p_1-p_2} = \mu_{p_1} - \mu_{p_2} = p_1 - p_2$$

و

$$\sigma^2_{p_1-p_2} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

مثال (7.7): إذا علمت أن 4% من السيارات المُصنعة في الشركة 1 لونها أحمر، و 3% من السيارات المُصنعة في الشركة 2 لونها أيضا أحمر. وتم سحب عينة حجمها 400 سيارة من كل شركة. وعلى افتراض أن الفرق بين نسب السيارات المنتجة في الشركتين التي لونها أحمر يتبع توزيع طبيعي، فأوجد:

1. الوسط الحسابي والتباين لتوزيع المعاينة للفرق بين نسب السيارات الحمراء المُصنعة في الشركتين.
2. احتمال أن تتعدى نسبة السيارات الحمراء المنتجة في الشركة 1 نسبة السيارات الحمراء المنتجة في الشركة 2 بمقدار 3% على الأقل.

الحل:

1. لدينا

$$\mu_{p_1-p_2} = \mu_{p_1} - \mu_{p_2} = p_1 - p_2 = 0.04 - 0.03 = 0.01$$

وكذلك

$$\sigma^2_{p_1-p_2} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} = \frac{0.04(0.96)}{400} + \frac{0.03(0.97)}{400} = 0.00017$$

2. حيث أن توزيع المعاينة للفرق بين نسبتي الشركتين يتبع التوزيع الطبيعي، نستطيع كتابة

$$P(P_1 - P_2 \geq 0.03) = P\left(\frac{(P_1 - P_2) - \mu_{p_1-p_2}}{\sigma_{p_1-p_2}} \geq \frac{0.03 - \mu_{p_1-p_2}}{\sigma_{p_1-p_2}}\right)$$

وبالتالي

$$P(P_1 - P_2 \geq 0.03) = P\left(Z \geq \frac{0.03 - 0.01}{0.013}\right) = P(Z \geq 1.54) = 1 - P(Z < 1.54)$$
$$= 1 - 0.938 = 0.062$$

وفيما تبقى من هذا الفصل، سنقوم بتعريف بعض أهم توزيعات المعاينة الاحتمالية والتي، إلى جانب التوزيع الطبيعي، تشكل الأساس الاحتمالي الذي تعتمد عليه نظريات وأساليب الاستدلال الإحصائي.

4.7 توزيع استيودنت t (Student's t Distribution)

في نظرية النهاية المركزية (نظرية (3.7)) رأينا أن المتغير العشوائي $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ يتوزع بتوزيع طبيعي معياري، حيث كان للمتغير $\bar{X}$ توزيع معاينة يقترب من التوزيع الطبيعي بمتوسط μ وتباين σ^2/n ، (علما بأن μ و σ^2 هما المتوسط والتباين للمجتمع الأصلي).

الآن نريد طرح التساؤل التالي:

ماذا يحدث لتوزيع المعاينة للمتغير $\bar{X}$ ، (وبالتالي توزيع المتغير Z)، إذا كانت قيمة تباين المجتمع σ^2 مجهولة لسبب ما ؟. الإجابة ستكون أننا سنقف أمام حالتين:

الحالة الأولى عندما يكون حجم العينة المسحوبة من المجتمع كبير[1]، وعند ذلك يمكن استخدام بديل (تقدير) لهذا التباين وهو عبارة عن تباين العينة (ويرمز له عادة بالرمز S^2). وطالما أن تباين العينة يعبر عن تباين المجتمع بشكل جيد ولا تتغير قيمته كثيرا من عينة لأخرى فإن المتغير العشوائي $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ سيظل يتبع توزيعا طبيعيا معياريا تقريبا بموجب نص نظرية النهاية المركزية.

أما في الحالة الثانية، والتي يكون فيها حجم العينة المسحوبة صغيرا، ($n < 30$)، فإن قيمة تباين العينة S^2 ستتغير بصورة كبيرة من عينة لأخرى، وبالتالي فإن توزيع المتغير $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ لن يتبع توزيعا طبيعيا معياريا، بل سيتبع توزيعا آخر يعرف بتوزيع[2] استيودنت t ، أو توزيع t . ويعتمد هذا التوزيع على معلمة واحدة هي حجم العينة ناقصا واحد، ($n - 1$) ، وهو ما يعرف **بدرجات الحرية**[3] (Degrees of Freedom) لتوزيع t .

نظرية (5.7): إذا كانت $\bar{X}$ و S^2 هما، على الترتيب، الوسط الحسابي والتباين لعينة عشوائية حجمها n مسحوبة من مجتمع يتوزع بتوزيع طبيعي متوسطه μ وتباينه σ^2 فإن المتغير العشوائي

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

سيتبع توزيع t بدرجات حرية $\gamma = n - 1$ ، حيث γ هي معلمة التوزيع.

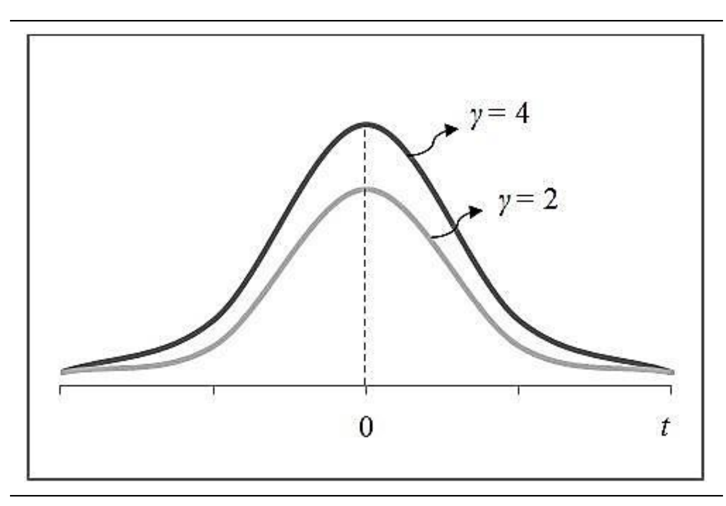
ويأخذ منحنى توزيع t شكلا ناقوسيا يشبه كثيرا منحنى التوزيع الطبيعي. وتتغير درجة تفرطح المنحنى بتغير قيمة معلمة التوزيع والتي تعتمد كما هو واضح على حجم العينة n ، كما يشاهد من الشكل (2.7). ويؤول التوزيع إلى الشكل الطبيعي المعياري كلما زادت قيمة n ، ($n \to \infty$).

---

[1] تُشير الكثير من الكتب الإحصائية إلى أن حجم العينة يعتبر كبيرا إذا تعدى 29 مفردة، ($n \geq 30$).

[2] قام باشتقاق هذا التوزيع العالم و. س. جوسيه (W. S. Gosset)، وقد سمي هذا التوزيع باسم استيودنت لأنه نشر بذلك الاسم.

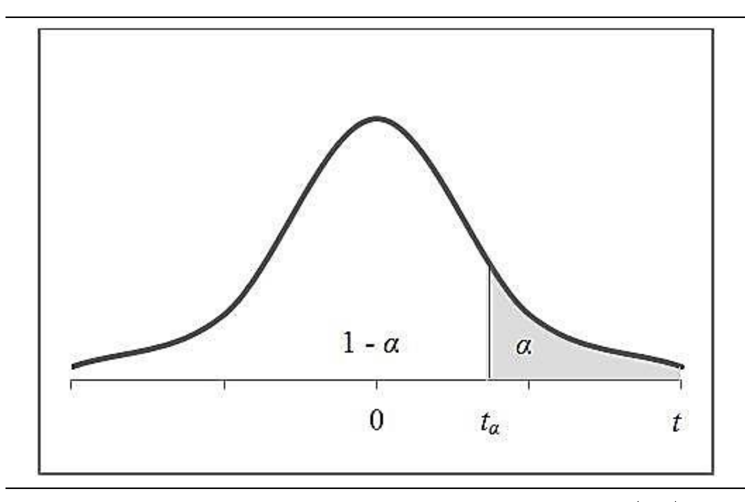
[3] تشير **درجات الحرية** هنا إلى عدد مفردات البيانات (العينة) التي تكون "حرة" في تغيرها بعد حساب قيمة الوسط الحسابي للعينة. إلا أن مفهوم درجات الحرية له تعريف أكثر عمقا ضمن إطار الاستدلال الإحصائي.



شكل (2.7): منحنى توزيع t بدرجات حرية مختلفة.

ويلاحظ من شكل منحنى التوزيع أنه كلما زادت قيم درجات الحرية (أي معلمة التوزيع) فإن المنحنى يصبح مدببا أكثر، أي يقل التشتت، وهذا يعني أن منحنى التوزيع الطبيعي هو أقل تشتتا (له تباين أقل) من منحنى توزيع t ، علما بأن مجموع المساحة الكلية تحت منحنى توزيع t يساوي الواحد الصحيح.

وبالتالي إذا أردنا حساب احتمال أن يكون المتغير T أكبر من قيمة معينة t_α مثلا فهذا يعني أن المطلوب هو حساب الاحتمال[1] $P(T > t_\alpha) = \alpha$ كما يتضح من الشكل (3.7). وبالتالي يكون $P(T < t_\alpha) = 1 - \alpha$.



شكل (3.7): حساب المساحة تحت منحنى توزيع t والتي تمثل الاحتمال $P(T > t_\alpha)$.

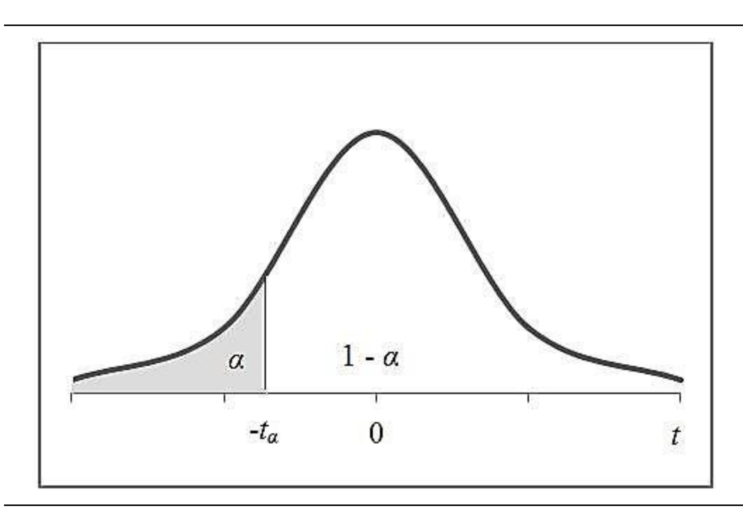
وحيث أن منحنى توزيع t هو منحنى متماثل ويكون في النصف الأيسر منه قيم T السالبة فإن الاحتمالات في هذا النصف يمكن حسابها بالصورة $P(T < -t_\alpha) = \alpha$ ، وهذا ما يوضحه الشكل (4.7). أي أنه يمكن كتابة أن:

$$P(T > t_\alpha) = P(T < -t_\alpha) = \alpha$$

أو

$$t_{1-\alpha} = - t_\alpha$$

[1] هذا الاحتمال يُقرأ كالتالي: احتمال أن تكون قيمة المتغير T أكبر من القيمة t_α يساوي α .



شكل (4.7): حساب المساحة تحت منحنى توزيع t والتي تمثل الاحتمال $P(T < -t_\alpha)$.

تعريف (5.7): **دالة الكثافة الاحتمالية لتوزيع t** : إذا كان T متغير عشوائي يتبع توزيع t بدرجات حرية تساوي $\gamma = n - 1$ فإن دالة الكثافة الاحتمالية له تعرّف بالصورة:

$$f_T(t) = k\left(1 + \frac{t^2}{\gamma}\right)^{-(\gamma+1)/2} \quad , \qquad -\infty < t < \infty$$

حيث k هو ثابت تحدد قيمته بحيث يكون $\int_{-\infty}^{\infty} f_T(t)dt = 1$.

إلا أن دالة الكثافة الاحتمالية السابقة لا تُستخدم عمليا لحساب الاحتمالات الخاصة بالمتغير العشوائي، بل يتم الاعتماد على القيم الاحتمالية الجدولية لتوزيع t، (في ملحق 1)، كما هو الحال مع التوزيع الطبيعي المعياري.

مثال (8.7): أوجد:

1. $P(T < t_{0.05})$
2. $P(-t_{0.025} < T < t_{0.05})$

الحل:

1. $P(T < t_{0.05}) = 1 - P(T > t_{0.05}) = 1 - 0.05 = 0.95$
2. $P(-t_{0.025} < T < t_{0.05}) = P(T < t_{0.05}) - P(T < -t_{0.025}) = 0.95 - 0.025 = 0.925$

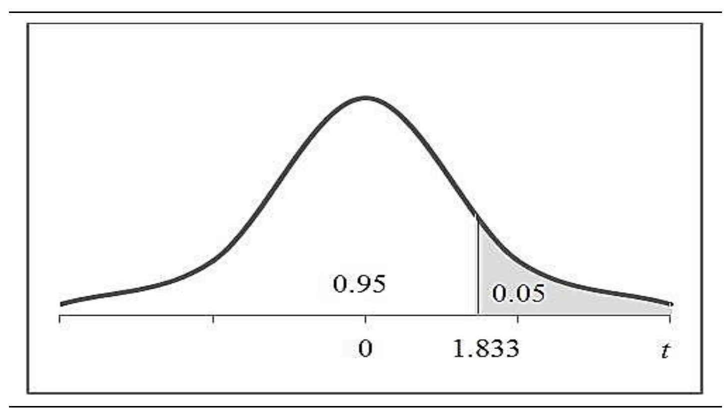
مثال (9.7): إذا كان المتغير العشوائي T يتبع توزيع t بمعلمة γ فأوجد القيم التالية:

1. $t_{0.05}$ إذا كانت $n = 10$
2. $t_{0.025}$ إذا كانت $n = 5$
3. $t_{0.99}$ إذا كانت $n = 21$

الحل:

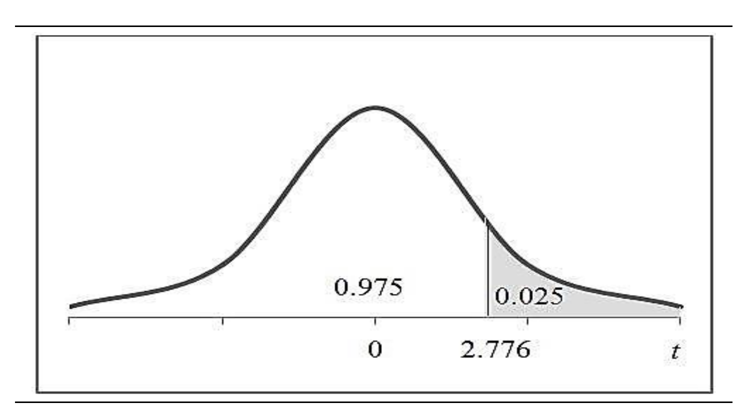
1. لدينا $\gamma = n - 1 = 10 - 1 = 9$ فيكون المطلوب إيجاد $t_{0.05}(9)$. من جدول t (ملحق 1، جدول (م3))، نقوم أولا بتحديد قيمة α حيث يندرج تحت كل قيمة لـ α عامود به قيم تتغير بتغير قيم درجات الحرية γ ، ثم نقوم بتحديد العامود الذي يحوي القيمة $\alpha = 0.05$ ثم نبحث عن درجة الحرية $\gamma = 9$ في العامود الأول من

اليسار، فتكون قيمة t هي القيمة المناظرة لـ $\alpha = 0.05$ و $\gamma = 9$ وهي $t_{0.05}(9) = 1.833$ ، وهذه القيمة يمكن ملاحظتها أيضا من منحنى توزيع t في الشكل (5.7).



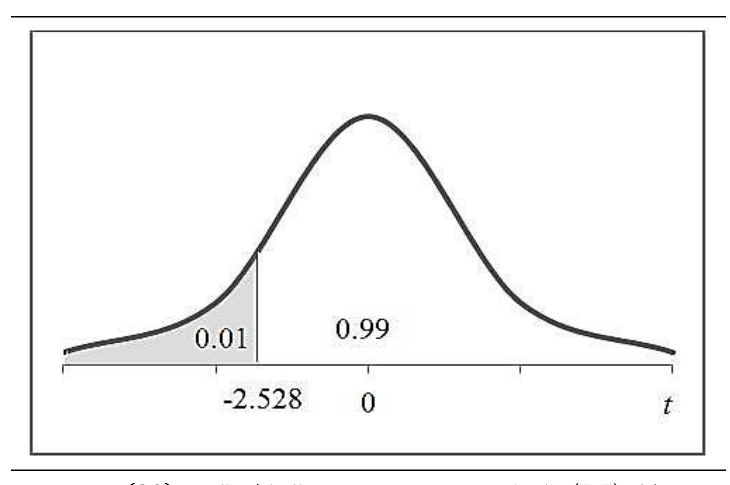
شكل (5.7): المساحة تحت منحنى توزيع t المناظرة للقيمة $t_{0.05}(9)$.

2. في جدول t من جديد في العامود تحت القيمة $\alpha = 0.025$ نجد مقابل القيمة $\gamma = 5$ -1 =4 أن $t_{0.025}(4) = 2.776$. كما يوضح الشكل (6.7).



شكل (6.7): المساحة تحت منحنى توزيع t المناظرة للقيمة $t_{0.025}(4)$.

3. من الشكل (7.7) نجد أن قيمة t المطلوبة هي التي تقسم منحنى التوزيع بحيث يكون على يمين القيمة $t_{0.99}(20)$ المساحة 0.99 وعلى يسارها المساحة 0.01 ، وبالتالي فإن $t_{0.99}(20) = -t_{0.01}(20)$.



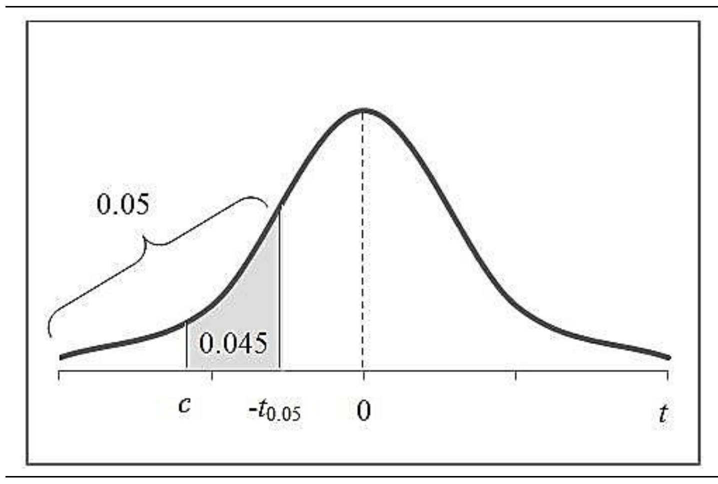
شكل (7.7): المساحة تحت منحنى توزيع t المناظرة للقيمة $t_{0.99}(20)$.

من جدول t نجد أن $t_{0.01}(20) = 2.528$ وهكذا فإن $t_{0.99}(20) = -2.528$.

مثال (10.7): إذا تم سحب عينة عشوائية حجمها $n = 15$ من مجتمع يتوزع بتوزيع طبيعي فأوجد قيمة الثابت c التي تحقق الاحتمال $P(c < T < -1.761) = 0.045$.

الحل:

نبدأ بالبحث عن القيمة 1.761 بداخل جدول t فنجد أنها تناظر القيمة $\alpha = 0.05$ وقيمة درجة الحرية $\gamma = 14$ ، أي أن $-t_{0.05}(14) = -1.761$ ، وحيث أن قيمة الثابت c هي أقل من -1.761 فستكون هي الأخرى سالبة ويكون الاحتمال المطلوب كما هو موضح في الشكل (8.7)، وهذا يعني أن المساحة على يسار القيمة $-t_{0.05}$ تساوي 0.05 .



شكل (8.7): المساحة تحت منحنى توزيع t المناظرة للمثال (10.7).

وحيث أن

$$P(c < T < -1.761) = P(T < -1.761) - P(T < c)$$

$$0.045 = 0.05 - P(T < c)$$

$$P(T < c) = 0.005$$

فإنه من جدول t نجد أن قيمة c التي تحقق $P(T < c) = 0.005$ ، (أو $-t_{0.005}(14)$) هي -2.977 ، أي أن $P(-2.977 < T < -1.761) = 0.045$.

تعريف (6.7): **التوقع والتباين لتوزيع t:** إذا كان T متغير عشوائي يتبع توزيع t بمعلمة $\gamma = (n - 1)$ فإن توقعه وتباينه يعطى بالصورة التالية:

$$E(T) = 0 \quad , \quad Var(T) = \frac{\gamma}{\gamma - 2} \quad , \quad \gamma > 2$$

وهذا يعني أنه لا يمكن حساب التباين لتوزيع t إذا كان حجم العينة $n < 3$.

5.7 توزيع مربع كاي (Chi-Square Distribution)

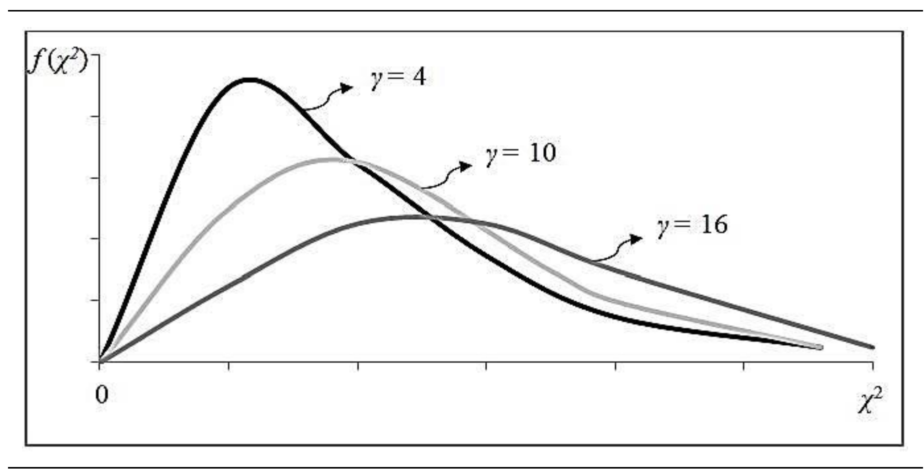
إن توزيع مربع كاي[1] يعد أيضا من توزيعات المعاينة الشهيرة التي ترتبط بالتوزيع الطبيعي، وهو توزيع متصل موجب غير متماثل لا يأخذ فيه المتغير العشوائي قيما سالبة لأن قيمه يتم حسابها من خلال صيغة موجبة كما تنص النظرية التالية:

نظرية (6.7): إذا كان S^2 هو تباين عينة عشوائية حجمها n مسحوبة من مجتمع يتوزع بتوزيع طبيعي له التباين σ^2 فإن:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

هو متغير عشوائي يتبع توزيع مربع كاي بدرجات حرية $\gamma = n - 1$ والتي تعد معلمة التوزيع.

وكما هو الحال مع منحنى توزيع t فإن منحنى توزيع χ^2 يتغير شكله بتغير معلمة التوزيع γ كما يتضح من الشكل (9.7).



شكل (9.7): منحنى توزيع χ^2 لقيم γ المختلفة.

ويتم حساب الاحتمالات المناظرة لقيم χ^2 المختلفة باستخدام جدول مربع كاي (في ملحق 1) بحيث يكون

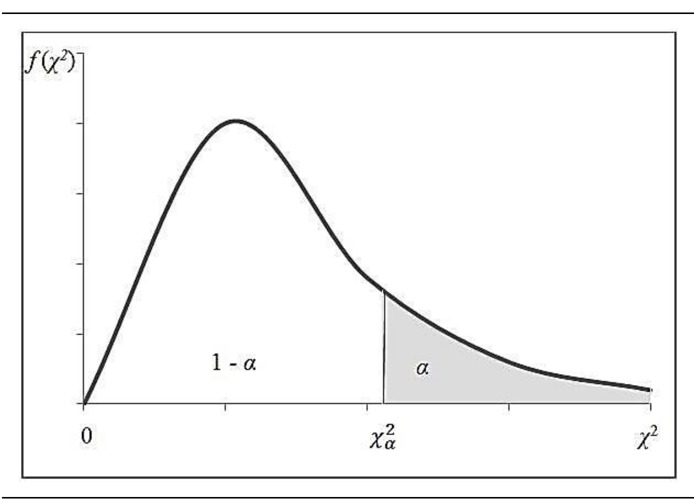
$$P(\chi^2 \geq \chi^2_\alpha) = \alpha$$

و

$$P(\chi^2 < \chi^2_\alpha) = 1 - \alpha$$

كما يوضح الشكل (10.7).

[1] كاي (Chi) هو اسم الحرف اليوناني الذي يرمز له بالرمز χ .



شكل (10.7): المساحة (الاحتمال) α تحت منحنى توزيع χ^2.

تعريف (7.7): **دالة الكثافة الاحتمالية لتوزيع χ^2** : إذا كان X^2 متغير عشوائي[1] يتبع توزيع مربع كاي بدرجات حرية γ فإن دالة كثافته الاحتمالية تعرّف بالصورة:

$$f_{X^2}(\chi^2) = k\,(\chi^2)^{\frac{\gamma}{2}-1}\, e^{-\chi^2/2} \quad , \quad 0 \le \chi^2 \le \infty$$

و k ثابت تحدد قيمته بحيث يكون $\int_0^\infty f_{X^2}(\chi^2)d\chi^2 = 1$.

ملاحظة: لتبسيط التعامل مع دالة الكثافة الاحتمالية لتوزيع مربع كاي يمكن التعويض عن قيمة المتغير χ^2 بالرمز المعتاد لقيمة المتغير العشوائي x ، عندئذ يمكن كتابة الدالة السابقة (في تعريف (7.7)) بالصورة:

$$f(x) = k\,(x)^{\frac{\gamma}{2}-1}\, e^{-x/2} \quad , \quad 0 \le x < \infty$$

وهي أقرب للصورة الاعتيادية في كتابة التوزيعات الاحتمالية.

وتوجد علاقات تربط توزيع مربع كاي بتوزيعات أخرى كما سنرى من التعريفات التالية.

تعريف (8.7): **توزيع χ^2 كحالة خاصة من توزيع جاما:** يقال أن المتغير العشوائي X يتبع توزيع مربع كاي بدرجات حرية γ إذا وفقط إذا كان X يتبع توزيع جاما بالمعالم $\gamma/2$ و 2 ، حيث γ هو عدد موجب.

تعريف (9.7): **علاقة توزيع t بتوزيع مربع كاي والتوزيع الطبيعي المعياري**: إذا كان Z متغير عشوائي يتبع التوزيع الطبيعي المعياري، وكان χ^2 متغير عشوائي يتبع توزيع مربع كاي بدرجات حرية γ ، عندئذ إذا كان Z و χ^2 مستقلان فإن:

$$T = \frac{Z}{\sqrt{\chi^2/\gamma}}$$

هو متغير عشوائي يتبع توزيع t بدرجات حرية γ .

---

[1] الرمز X هو الحرف الكبير لـ كاي والذي يشبه الحرف "إكس" في اللغة الإنجليزية. والقيمة 2 في أُس المتغير X^2 هي للدلالة على تسمية المتغير (مربع كاي) وليست قيمة تربيعية.

تعريف (10.7): **علاقة توزيع مربع كاي بالتوزيع الطبيعي المعياري:** إذا كانت $X_1, X_2, \dots , X_m$ هي متغيرات عشوائية مستقلة وكل منها يتبع التوزيع الطبيعي المعياري فإن مجموع مربعات هذه المتغيرات، $\sum_{i=1}^{m} X_i^2$ يتبع توزيع مربع كاي بدرجات حرية m .

مثال (11.7): أوجد قيم المتغير العشوائي χ^2 التالية:

1. $\chi^2_{0.025}$ إذا كان $n = 12$ 2. $\chi^2_{0.99}$ إذا كان $n = 26$

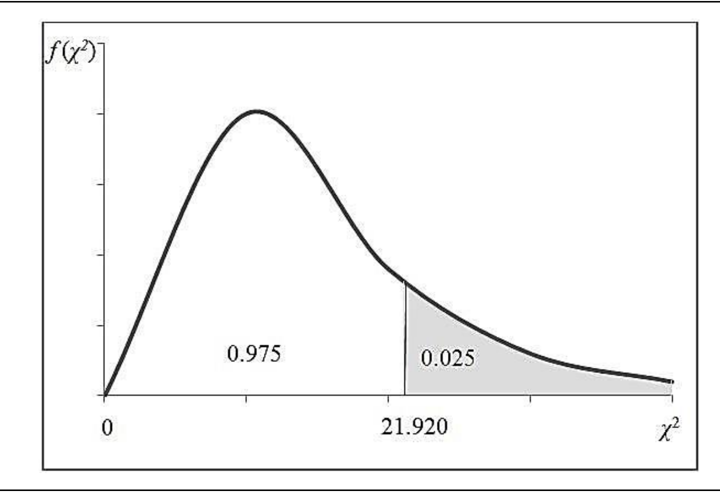
الحل:

1. البحث في جدول مربع كاي، (ملحق 1، جدول (م4))، يشبه البحث في جدول t ، فعندما نريد إيجاد القيمة $\chi^2_{0.025}$ عندما $n = 12$ فهذا يعني البحث عن $\chi^2_{0.025}(11)$ ، وهكذا فإننا نبحث عن القيمة التي تناظر $\alpha = 0.025$ و $\gamma = 11$ والتي سنجد أنها 21.920 كما يوضح الشكل (11.7). ونلاحظ أن

$$P(\chi^2 > \chi^2_\alpha) = P(\chi^2 > 21.920) = 0.025 = \alpha$$

و

$$P(\chi^2 < \chi^2_\alpha) = P(\chi^2 < 21.920) = 0.975 = 1 - \alpha$$



شكل (11.7): المساحة تحت منحنى توزيع χ^2 للمثال (11.7) فقرة 1.

2. من جدول مربع كاي يكون $\chi^2_{0.99}(25) = 11.524$.

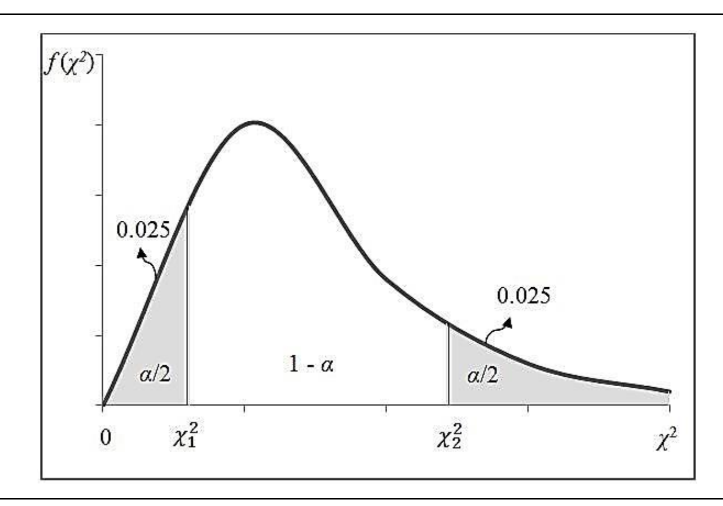
مثال (12.7): أوجد قيم χ^2 التي تحقق الاحتمال $P(\chi_1^2 < \chi^2 < \chi_2^2) = \alpha$ علما بأن $\alpha = 0.05$ ، و $n = 6$ وأن المساحة على يمين القيمة χ_2^2 تساوي المساحة على يسار القيمة χ_1^2.

الحل:

من الشكل (12.7) الذي يوضح معطيات المثال، نلاحظ أن قيمة χ_1^2 التي تحقق

$$P(\chi^2 < \chi_1^2) = 1 - 0.025$$

هي $\chi_1^2 = \chi^2_{0.975}(5) = 0.831$.



شكل (12.7): تقسيم المساحة تحت منحنى توزيع χ^2 للمثال (12.7).

وأن قيمة χ_2^2 التي تحقق

$$P(\chi^2 > \chi_2^2) = 0.025$$

هي $\chi_2^2 = \chi_{0.025}^2(5) = 12.832$.

تعريف (11.7): **التوقع والتباين لتوزيع مربع كاي:** إذا كان χ^2 متغير عشوائي يتبع توزيع مربع كاي بدرجات حرية γ فإن التوقع والتباين له يعرّف بالصورة:

$$E(\chi^2) = \gamma \quad , \quad Var(\chi^2) = 2\,\gamma$$

6.7 توزيع فيشر F (Fisher's F Distribution)

إن توزيع[1] F يُعتبر من التوزيعات الاحتمالية المتصلة الهامة خاصة في مجال تحليل التباين في الإحصاء الاستدلالي، حيث أنه يتعامل عادة مع مقاييس التباين لأكثر من مجتمع. إذا كان لدينا مجتمعين مستقلين كل منهما يتوزع بتوزيع مربع كاي (ولتكن χ_1^2 و χ_2^2) بدرجات حرية γ_1 للمجتمع الأول و γ_2 للمجتمع الثاني، فإن المتغير العشوائي F يمكن تعريفه نظريا بأنه النسبة بين χ_1^2 و χ_2^2 بعد قسمة كل منهما على درجات الحرية الخاصة به، بمعنى أن:

$$F = \frac{\chi_1^2/\gamma_1}{\chi_2^2/\gamma_2} = \frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}/(n_1-1)}{\frac{(n_2-1)S_2^2}{\sigma_2^2}/(n_2-1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \quad ; \quad S_1^2 > S_2^2$$

حيث أن (n_1 ، S_1^2 ، و σ_1^2) و (n_2 ، S_2^2 ، و σ_2^2) هي أحجام العينات المسحوبة من المجتمعين وتبايناتها و تباينات المجتمعين على الترتيب.

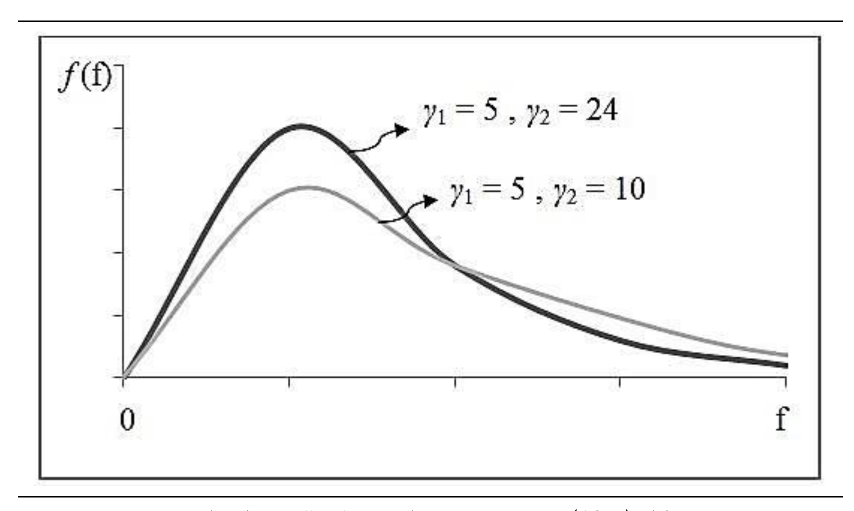
[1] والذي يُعرف أيضا بتوزيع **فيشر– سنيديكور** (Fisher-Snedecor) نسبة للعالمين رونالد فيشر وجورج سنيديكور.

نظرية (7.7): إذا كان S_1^2 و S_2^2 هما تبايني عينتين عشوائيتين مستقلتين، حجم الأولى n_1 والثانية n_2 ، ومسحوبتين من مجتمعين يتوزعان بتوزيع طبيعي بتباينات σ_1^2 و σ_2^2 على الترتيب، فإن

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \quad , \qquad S_1^2 > S_2^2$$

هو متغير عشوائي يتوزع بتوزيع F بدرجتي حرية $\gamma_1 = n_1 - 1$ و $\gamma_2 = n_2 - 1$.

ومنحنى توزيع F هو منحنى موجب يشبه منحنى توزيع مربع كاي، لأنه يعتمد عليه، كما يوضح الشكل (13.7).



شكل (13.7): منحنى توزيع F لقيم درجات الحرية المختلفة.

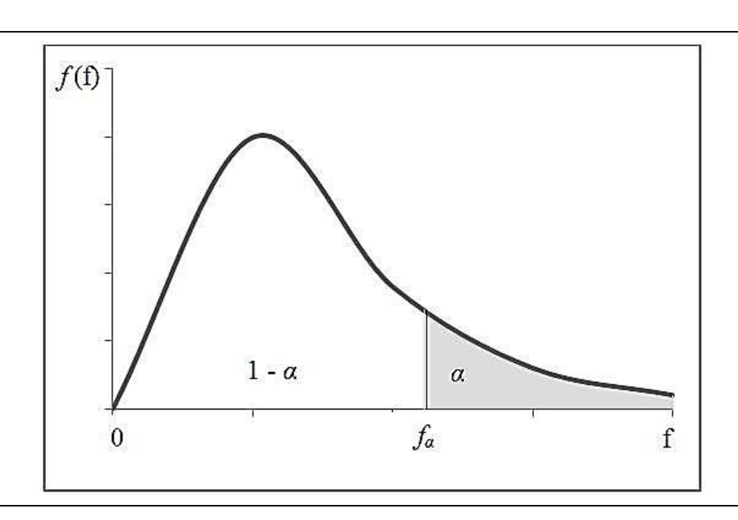
ويتم حساب الاحتمالات لقيم F المختلفة كما هو الحال مع توزيع مربع كاي أيضا حيث يكون

$$P(F \geq f_\alpha) = \alpha$$

و

$$P(F < f_\alpha) = 1 - \alpha$$

كما يتضح من الشكل (14.7).



شكل (14.7): المساحة تحت منحنى توزيع F المناظرة للاحتمال $P(F > f_\alpha) = \alpha$.

نظرية (8.7): إذا كانت $f_\alpha(\gamma_1,\gamma_2)$ هي قيمة المتغير العشوائي F بدرجات حرية γ_1 و γ_2 على الترتيب، والتي تكون مساحة المنحنى على يمينها تساوي α فإن:

$$f_{1-\alpha}(\gamma_1,\gamma_2)=\frac{1}{f_\alpha(\gamma_2,\gamma_1)}$$

إن النظرية (7.7) توضح لنا العلاقة بين توزيع F وتوزيع مربع كاي، أما الشكل الرياضي للتوزيع فيقدمه لنا التعريف التالي.

تعريف (12.7): إذا كان F متغير عشوائي يتبع توزيع فيشر F بالمعالم γ_1 و γ_2 فإن دالة الكثافة الاحتمالية له تعرف بالصورة:

$$f_F(\mathrm{f})=k\ (\mathrm{f})^{\frac{\gamma_1}{2}-1}\ (\gamma_2+\gamma_1\mathrm{f})^{-(\gamma_1+\gamma_2)/2}\quad ,\quad 0\le \mathrm{f}<\infty$$

حيث أن k هو مقدار ثابت.

وتوزيع F تربطه علاقة رياضية أيضا بتوزيع t توضحها النظرية التالية.

نظرية (9.7): إذا كان F متغير عشوائي يتبع توزيع فيشر F بدرجات حرية 1 و γ_2 ، فإن:

$$f_\alpha(1,\gamma_2)=t^2_{\alpha/2}(\gamma_2)$$

بمعنى أن قيمة المتغير F بدرجات حرية 1 و γ_2 تكافئ مربع قيمة المتغير T الذي يتبع توزيع t بدرجة حرية γ_2 ، ويقسم مساحة منحنى t إلى $\alpha/2$ إلى اليمين.

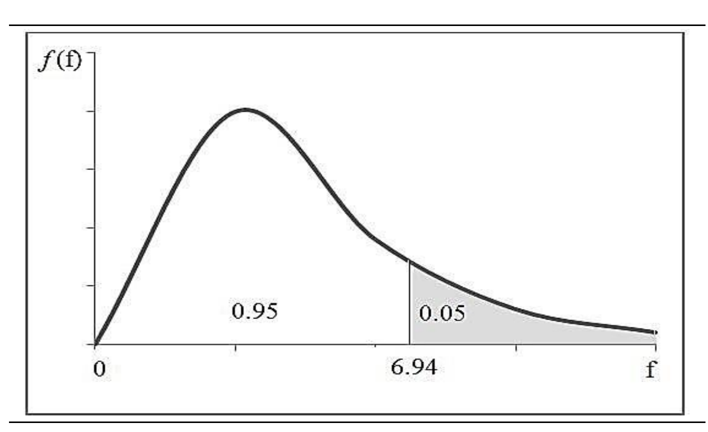
مثال (13.7): أوجد قيم F التالية:

1. $f_{0.05}$ إذا كانت $n_1 = 3$ و $n_2 = 5$
2. $f_{0.01}$ إذا كانت $n_1 = n_2 = 11$
3. $f_{0.10}$ إذا كانت $n_1 = 2$ و $n_2 = 7$

الحل:

1. سنستخدم جدول F (ملحق 1، جدول (م5)) الذي يحتوي على خمسة قيم لـ α هي 0.10، 0.05، 0.01، 0.025، و 0.001 .

نبدأ بتحديد الجدول الذي يحتوي القيمة $\alpha = 0.05$ ، ثم نبحث عن القيمة $\gamma_1 = 2$ (لأن $n_1 = 3$) في العامود الأول الخاص بدرجة الحرية الأولى، والقيمة $\gamma_2 = 4$ (لأن $n_2 = 5$) في الصف الأول الخاص بدرجة الحرية الثانية، فتكون قيمة F المناظرة لهاتين الدرجتين هي $f_{0.05}(2{,}4) = 6.94$. والشكل (15.7) يوضح موقع هذه القيمة على منحنى توزيع F .



شكل (15.7): المساحة تحت منحنى توزيع F للقيمة $f_{0.05}(2,4)$.

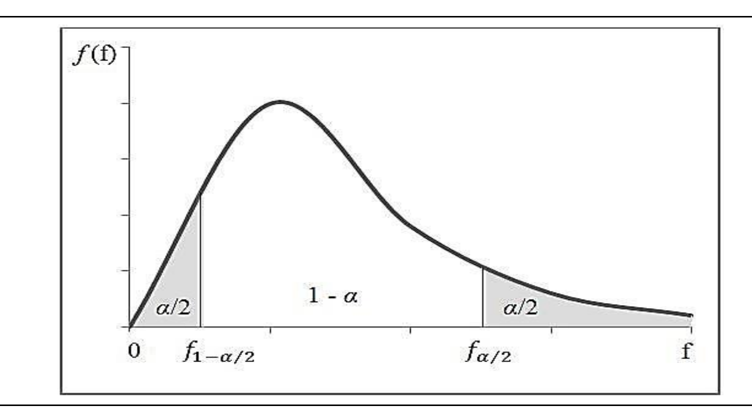
2. في جدول F ذو القيمة $\alpha = 0.01$ نجد أن $f_{0.01}(10,10) = 4.85$.

3. يمكن في هذه الحالة استخدام جدول F الذي يحتوي عدة قيم لـ α من ضمنها $\alpha = 0.10$ ، أو ببساطة يمكن استخدام العلاقة بين توزيع F و توزيع t كما في النظرية (9.7) لأن $\gamma_1 = 1$ ، وهكذا يكون لدينا

$$f_\alpha(1,\gamma_2) = f_{0.10}(1,6) = t^2_{\alpha/2}(\gamma_2) = t^2_{0.10/2}(6) = t^2_{0.05}(6) = (1.943)^2 = 3.775$$

ملاحظة: إذا ما تم تقسيم المنطقة المظللة تحت منحنى توزيع F إلى مساحتين متساويتين في الأطراف، كما في الشكل (16.7)، فإن احتمال أن تكون قيم المتغير F محصورة بين القيمتين $f_{\alpha/2}$ و $f_{1-\alpha/2}$ يساوي

$$P\left(f_{1-\alpha/2}(\gamma_1,\gamma_2) < F < f_{\alpha/2}(\gamma_1,\gamma_2)\right) = (1-\alpha)$$



شكل (16.7): تقسيم المساحة تحت منحنى توزيع F إلى مساحتين متساويتين في الأطراف.

تعريف (13.7): **التوقع والتباين لتوزيع F** : إذا كان F متغير عشوائي يتبع توزيع فيشر F بدرجات حرية γ_1 و γ_2 فإن توقعه وتباينه يعرّف بالصورة

$$E(F) = \frac{\gamma_2}{\gamma_2 - 2} \quad , \quad \gamma_2 > 2$$

و

$$Var(F) = \frac{2\,{\gamma_2}^2\,(\gamma_1 + \gamma_2 - 2)}{\gamma_1(\gamma_2 - 2)^2\,(\gamma_2 - 4)} \quad , \quad \gamma_2 > 4$$

7.7 تمارين الفصل السابع

تمرين (1.7): أوجد توزيع المعاينة للوسط $\bar{X}$ إذا ما تم سحب كل العينات الممكنة بالإرجاع ذات الحجم $n = 3$ من المجتمع الذي يتكون من البيانات 3، 5، 6 . وكذلك أوجد الوسط الحسابي والتباين لتوزيع المعاينة للوسط.

تمرين (2.7): في التمرين (1.7) أوجد توزيع المعاينة للوسط إذا ما تم سحب كل العينات الممكنة بدون إرجاع ذات الحجم $n = 2$ و $n = 3$. وكذلك أوجد الوسط الحسابي والتباين لتوزيع المعاينة للوسط.

تمرين (3.7): إذا ما تم اعتبار أن معدلات ذكاء الأطفال في مرحلة الدراسة الابتدائية تتبع التوزيع الطبيعي بمتوسط 5 درجات وانحراف معياري 1.5 درجة، وتم سحب كل العينات الممكنة ذات الحجم 20 طفل بالإرجاع من مجتمع الأطفال الافتراضي الذي حجمه 2000 طفل، فأوجد:

1. الوسط الحسابي والانحراف المعياري لتوزيع المعاينة للوسط $\bar{X}$ الناتج عن سحب العينات بالإرجاع.
2. احتمال سحب عينات لها أوساط أقل من 4 درجات.
3. احتمال سحب عينات لها أوساط أكبر من 5.5 درجات.
4. احتمال سحب عينات لها أوساط تقع بين 4 و 5.5 درجات.

تمرين (4.7): مستخدما مفردات المجتمعين التاليين أوجد:

المجتمع (1): 4، 6، 7	المجتمع (2): 2، 5، 6

1. متوسط وتباين كل مجتمع.
2. كل العينات الممكن سحبها بالإرجاع من المجتمع الأول ذات الحجم $n_1 = 2$ ، ومن المجتمع الثاني ذات الحجم $n_2 = 2$.
3. توزيع المعاينة للفرق بين أوساط العينات.
4. أثبت عدديا أن $\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2$ و $\sigma^2_{\bar{X}_1-\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$.

تمرين (5.7): أحد المصانع الكورية ينتج نوعان من السيارات، النوع LS والذي له متوسط استهلاك وقود يقدر بـ 70 لتر/500 كم بتباين 8 لتر/500 كم ، والنوع GTX والذي له متوسط استهلاك وقود يقدر بـ 50 لتر/500 كم بتباين 6 لتر/500 كم. وتم اختيار عينتين عشوائيتين حجم كل منهما 100 سيارة من كل نوع وتم اختبارها. ما احتمال أن السيارات من النوع LS سيكون لها متوسط استهلاك وقود أقل من النوع GTX بـ 19 لتر/500 كم ؟.

تمرين (6.7): إذا علمت أن نسبة مرضى اللوكيميا في المستشفيات الخاصة في إحدى العواصم الكبيرة هو 3%. فما هو احتمال أن يوجد في عينة مكونة من 300 مريض من تلك العاصمة:

1. أقل من 5% مريض باللوكيميا؟.
2. أكثر من 3% مريض باللوكيميا؟.

تمرين (7.7): في إحدى المدن، كانت نسبة الطلبة اللذين يعانون من عسر القراءة (Dyslexia) هو 5% في المدارس الحكومية و 3% في المدارس الخاصة. وتم اختيار عينة مكونة من 350 طالب من كلا من المدارس الحكومية والخاصة، فإذا علمت أن الفرق بين نسب الطلبة اللذين يعانون من عسر القراءة في نوعي المدارس يتبع التوزيع الطبيعي، فأوجد:

1. الوسط الحسابي والتباين لتوزيع المعاينة للفرق بين نسب الطلبة اللذين يعانون من عسر القراءة في نوعي المدارس.
2. احتمال أن لا تتعدى نسبة الطلبة اللذين يعانون من عسر القراءة في المدارس الحكومية نسبة الطلبة اللذين يعانون من عسر القراءة في المدارس الخاصة بمقدار 4%.

تمرين (8.7): إذا كان المتغير العشوائي T يتبع توزيع t بدرجات حرية γ فأوجد القيم التالية؛

1. $t_{0.005}$ إذا كانت $n = 15$ .2. $t_{0.01}$ إذا كانت $n = 7$ .3. $t_{0.95}$ إذا كانت $n = 25$.

تمرين (9.7): أوجد قيم المتغير العشوائي الذي يتبع توزيع مربع كاي التالية:

1. $\chi^2_{0.05}$ إذا كان $n = 15$ .2. $\chi^2_{0.01}$ إذا كان $n = 29$ .3. $\chi^2_{0.95}$ إذا كان $n = 10$.

تمرين (10.7): أوجد قيم المتغير العشوائي F التالية:

1. $f_{0.05}$ إذا كانت $n_1 = 6$ و $n_2 = 10$.
2. $f_{0.01}$ إذا كانت $n_1 = 9$ و $n_2 = 7$.
3. $f_{0.005}$ إذا كانت $n_1 = 4$ و $n_2 = 8$.

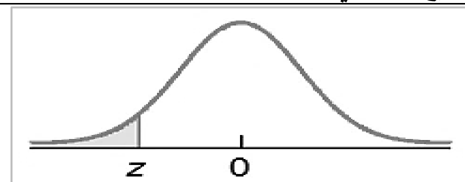
ملحق 1

الجداول الإحصائية (Statistical Tables)

جدول (م1): الأرقام العشوائية

رقم الصف	رقم العامود				
	01–10	**11–20**	**21–30**	**31–40**	**41–50**
01	**15544 80712**	**97742 21500**	**97081 42451**	**50623 56071**	**28882 28739**
02	**01011 21285**	**04729 39986**	**73150 31548**	**30168 76189**	**56996 19210**
03	**47435 53308**	**40718 29050**	**74858 64517**	**93573 51058**	**68501 42723**
04	**91312 75137**	**86274 59834**	**69844 19853**	**06917 17413**	**44474 86530**
05	**12775 08768**	**80791 16298**	**22934 09630**	**98862 39746**	**64623 32768**
06	**31466 43761**	**94872 92230**	**52367 13205**	**38634 55882**	**77518 36252**
07	**09300 43847**	**40881 51243**	**97810 18903**	**53914 31688**	**06220 40422**
08	**73582 13810**	**57784 72454**	**68997 72229**	**30340 08844**	**53924 89630**
09	**11092 81392**	**58189 22697**	**41063 09451**	**09789 00637**	**06450 85990**
10	**93322 98567**	**00116 35605**	**66790 52965**	**62877 21740**	**56476 49296**
11	**80134 12484**	**67089 08674**	**70753 90959**	**45842 59844**	**45214 36505**
12	**97888 31797**	**95037 84400**	**76041 96668**	**75920 68482**	**56855 97417**
13	**92612 27082**	**59459 69380**	**98654 20407**	**88151 56263**	**27126 63797**
14	**72744 45586**	**43279 44218**	**83638 05422**	**00995 70217**	**78925 39097**
15	**96256 70653**	**45285 26293**	**78305 80252**	**03625 40159**	**68760 84716**
16	**07851 47452**	**66742 83331**	**54701 06573**	**98169 37499**	**67756 68301**
17	**25594 41552**	**96475 56151**	**02089 33748**	**65289 89956**	**89559 33687**
18	**65358 15155**	**59374 80940**	**03411 94656**	**69440 47156**	**77115 99463**
19	**09402 31008**	**53424 21928**	**02198 61201**	**02457 87214**	**59750 51330**
20	**97424 90765**	**01634 37328**	**41243 33564**	**17884 94747**	**93650 77668**

جدول (م2): جدول التوزيع الطبيعي المعياري

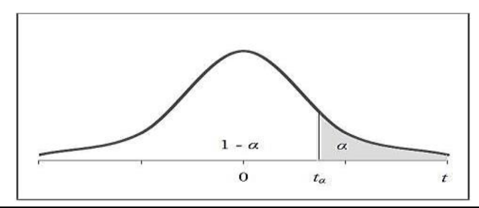


	α									
Z	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
-3.4	**.0003**	**.0003**	**.0003**	**.0003**	**.0003**	**.0003**	**.0003**	**.0003**	**.0003**	**.0002**
-3.3	**.0005**	**.0005**	**.0005**	**.0004**	**.0004**	**.0004**	**.0004**	**.0004**	**.0004**	**.0003**
-3.2	**.0007**	**.0007**	**.0006**	**.0006**	**.0006**	**.0006**	**.0006**	**.0005**	**.0005**	**.0005**
-3.1	**.0010**	**.0009**	**.0009**	**.0009**	**.0008**	**.0008**	**.0008**	**.0008**	**.0007**	**.0007**
-3.0	**.0013**	**.0013**	**.0013**	**.0012**	**.0012**	**.0011**	**.0011**	**.0011**	**.0010**	**.0010**
-2.9	**.0019**	**.0018**	**.0018**	**.0017**	**.0016**	**.0016**	**.0015**	**.0015**	**.0014**	**.0014**
-2.8	**.0026**	**.0025**	**.0024**	**.0023**	**.0023**	**.0022**	**.0021**	**.0021**	**.0020**	**.0019**
-2.7	**.0035**	**.0034**	**.0033**	**.0032**	**.0031**	**.0030**	**.0029**	**.0028**	**.0027**	**.0026**
-2.6	**.0047**	**.0045**	**.0044**	**.0043**	**.0041**	**.0040**	**.0039**	**.0038**	**.0037**	**.0036**
-2.5	**.0062**	**.0060**	**.0059**	**.0057**	**.0055**	**.0054**	**.0052**	**.0051**	**.0049**	**.0048**
-2.4	**.0082**	**.0080**	**.0078**	**.0075**	**.0073**	**.0071**	**.0069**	**.0068**	**.0066**	**.0064**
-2.3	**.0107**	**.0104**	**.0102**	**.0099**	**.0096**	**.0094**	**.0091**	**.0089**	**.0087**	**.0084**
-2.2	**.0139**	**.0136**	**.0132**	**.0129**	**.0125**	**.0122**	**.0119**	**.0116**	**.0113**	**.0110**
-2.1	**.0179**	**.0174**	**.0170**	**.0166**	**.0162**	**.0158**	**.0154**	**.0150**	**.0146**	**.0143**
-2.0	**.0228**	**.0222**	**.0217**	**.0212**	**.0207**	**.0202**	**.0197**	**.0192**	**.0188**	**.0183**
-1.9	**.0287**	**.0281**	**.0274**	**.0268**	**.0262**	**.0256**	**.0250**	**.0244**	**.0239**	**.0233**
-1.8	**.0359**	**.0351**	**.0344**	**.0336**	**.0329**	**.0322**	**.0314**	**.0307**	**.0301**	**.0294**
-1.7	**.0446**	**.0436**	**.0427**	**.0418**	**.0409**	**.0401**	**.0392**	**.0384**	**.0375**	**.0367**
-1.6	**.0548**	**.0537**	**.0526**	**.0516**	**.0505**	**.0495**	**.0485**	**.0475**	**.0465**	**.0455**
-1.5	**.0668**	**.0655**	**.0643**	**.0630**	**.0618**	**.0606**	**.0594**	**.0582**	**.0571**	**.0559**
-1.4	**.0808**	**.0793**	**.0778**	**.0764**	**.0749**	**.0735**	**.0721**	**.0708**	**.0694**	**.0681**
-1.3	**.0968**	**.0951**	**.0934**	**.0918**	**.0901**	**.0885**	**.0869**	**.0853**	**.0838**	**.0823**
-1.2	**.1151**	**.1131**	**.1112**	**.1093**	**.1075**	**.1056**	**.1038**	**.1020**	**.1003**	**.0985**
-1.1	**.1357**	**.1335**	**.1314**	**.1292**	**.1271**	**.1251**	**.1230**	**.1210**	**.1190**	**.1170**
-1.0	**.1587**	**.1562**	**.1539**	**.1515**	**.1492**	**.1469**	**.1446**	**.1423**	**.1401**	**.1379**
-0.9	**.1841**	**.1814**	**.1788**	**.1762**	**.1736**	**.1711**	**.1685**	**.1660**	**.1635**	**.1611**
-0.8	**.2119**	**.2090**	**.2061**	**.2033**	**.2005**	**.1977**	**.1949**	**.1922**	**.1894**	**.1867**
-0.7	**.2420**	**.2389**	**.2358**	**.2327**	**.2296**	**.2266**	**.2236**	**.2206**	**.2177**	**.2148**
-0.6	**.2743**	**.2709**	**.2676**	**.2643**	**.2611**	**.2578**	**.2546**	**.2514**	**.2483**	**.2451**
-0.5	**.3085**	**.3050**	**.3015**	**.2981**	**.2946**	**.2912**	**.2877**	**.2843**	**.2810**	**.2776**
-0.4	**.3446**	**.3409**	**.3372**	**.3336**	**.3300**	**.3264**	**.3228**	**.3192**	**.3156**	**.3121**
-0.3	**.3821**	**.3783**	**.3745**	**.3707**	**.3669**	**.3632**	**.3594**	**.3557**	**.3520**	**.3483**
-0.2	**.4207**	**.4168**	**.4129**	**.4090**	**.4052**	**.4013**	**.3974**	**.3936**	**.3897**	**.3859**
-0.1	**.4602**	**.4562**	**.4522**	**.4483**	**.4443**	**.4404**	**.4364**	**.4325**	**.4286**	**.4247**
0.0	**.5000**	**.4960**	**.4920**	**.4880**	**.4840**	**.4801**	**.4761**	**.4721**	**.4681**	**.4641**

تابع جدول (م2)

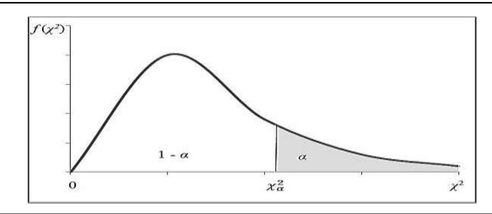
	α									
Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	**.5000**	**.5040**	**.5080**	**.5120**	**.5160**	**.5199**	**.5239**	**.5279**	**.5319**	**.5359**
0.1	**.5398**	**.5438**	**.5478**	**.5517**	**.5557**	**.5596**	**.5636**	**.5675**	**.5714**	**.5753**
0.2	**.5793**	**.5832**	**.5871**	**.5910**	**.5948**	**.5987**	**.6026**	**.6064**	**.6103**	**.6141**
0.3	**.6179**	**.6217**	**.6255**	**.6293**	**.6331**	**.6368**	**.6406**	**.6443**	**.6480**	**.6517**
0.4	**.6554**	**.6591**	**.6628**	**.6664**	**.6700**	**.6736**	**.6772**	**.6808**	**.6844**	**.6879**
0.5	**.6915**	**.6950**	**.6985**	**.7019**	**.7054**	**.7088**	**.7123**	**.7157**	**.7190**	**.7224**
0.6	**.7257**	**.7291**	**.7324**	**.7357**	**.7389**	**.7422**	**.7454**	**.7486**	**.7517**	**.7549**
0.7	**.7580**	**.7611**	**.7642**	**.7673**	**.7704**	**.7734**	**.7764**	**.7794**	**.7823**	**.7852**
0.8	**.7881**	**.7910**	**.7939**	**.7967**	**.7995**	**.8023**	**.8051**	**.8078**	**.8106**	**.8133**
0.9	**.8159**	**.8186**	**.8212**	**.8238**	**.8264**	**.8289**	**.8315**	**.8340**	**.8365**	**.8389**
1.0	**.8413**	**.8438**	**.8461**	**.8485**	**.8508**	**.8531**	**.8554**	**.8577**	**.8599**	**.8621**
1.1	**.8643**	**.8665**	**.8686**	**.8708**	**.8729**	**.8749**	**.8770**	**.8790**	**.8810**	**.8830**
1.2	**.8849**	**.8869**	**.8888**	**.8907**	**.8925**	**.8944**	**.8962**	**.8980**	**.8997**	**.9015**
1.3	**.9032**	**.9049**	**.9066**	**.9082**	**.9099**	**.9115**	**.9131**	**.9147**	**.9162**	**.9177**
1.4	**.9192**	**.9207**	**.9222**	**.9236**	**.9251**	**.9265**	**.9279**	**.9292**	**.9306**	**.9319**
1.5	**.9332**	**.9345**	**.9357**	**.9370**	**.9382**	**.9394**	**.9406**	**.9418**	**.9429**	**.9441**
1.6	**.9452**	**.9463**	**.9474**	**.9484**	**.9495**	**.9505**	**.9515**	**.9525**	**.9535**	**.9545**
1.7	**.9554**	**.9564**	**.9573**	**.9582**	**.9591**	**.9599**	**.9608**	**.9616**	**.9625**	**.9633**
1.8	**.9641**	**.9649**	**.9656**	**.9664**	**.9671**	**.9678**	**.9686**	**.9693**	**.9699**	**.9706**
1.9	**.9713**	**.9719**	**.9726**	**.9732**	**.9738**	**.9744**	**.9750**	**.9756**	**.9761**	**.9767**
2.0	**.9772**	**.9778**	**.9783**	**.9788**	**.9793**	**.9798**	**.9803**	**.9808**	**.9812**	**.9817**
2.1	**.9821**	**.9826**	**.9830**	**.9834**	**.9838**	**.9842**	**.9846**	**.9850**	**.9854**	**.9857**
2.2	**.9861**	**.9864**	**.9868**	**.9871**	**.9875**	**.9878**	**.9881**	**.9884**	**.9887**	**.9890**
2.3	**.9893**	**.9896**	**.9898**	**.9901**	**.9904**	**.9906**	**.9909**	**.9911**	**.9913**	**.9916**
2.4	**.9918**	**.9920**	**.9922**	**.9925**	**.9927**	**.9929**	**.9931**	**.9932**	**.9934**	**.9936**
2.5	**.9938**	**.9940**	**.9941**	**.9943**	**.9945**	**.9946**	**.9948**	**.9949**	**.9951**	**.9952**
2.6	**.9953**	**.9955**	**.9956**	**.9957**	**.9959**	**.9960**	**.9961**	**.9962**	**.9963**	**.9964**
2.7	**.9965**	**.9966**	**.9967**	**.9968**	**.9969**	**.9970**	**.9971**	**.9972**	**.9973**	**.9974**
2.8	**.9974**	**.9975**	**.9976**	**.9977**	**.9977**	**.9978**	**.9979**	**.9979**	**.9980**	**.9981**
2.9	**.9981**	**.9982**	**.9982**	**.9983**	**.9984**	**.9984**	**.9985**	**.9985**	**.9986**	**.9986**
3.0	**.9987**	**.9987**	**.9987**	**.9988**	**.9988**	**.9989**	**.9989**	**.9989**	**.9990**	**.9990**
3.1	**.9990**	**.9991**	**.9991**	**.9991**	**.9992**	**.9992**	**.9992**	**.9992**	**.9993**	**.9993**
3.2	**.9993**	**.9993**	**.9994**	**.9994**	**.9994**	**.9994**	**.9994**	**.9995**	**.9995**	**.9995**
3.3	**.9995**	**.9995**	**.9995**	**.9996**	**.9996**	**.9996**	**.9996**	**.9996**	**.9996**	**.9997**
3.4	**.9997**	**.9997**	**.9997**	**.9997**	**.9997**	**.9997**	**.9997**	**.9997**	**.9997**	**.9998**

جدول (م3): جدول توزيع استيودنت t



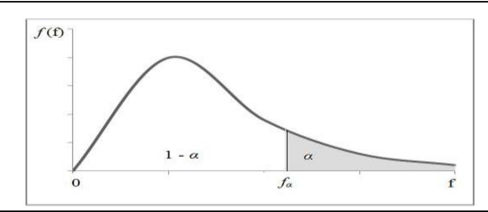
	α											
d.f	**.25**	**.20**	**.15**	**.10**	**.05**	**.025**	**.02**	**.01**	**.005**	**.0025**	**.001**	**.0005**
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
Z	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291

جدول (م4): جدول توزيع مربع كاي



	α											
d.f	**.25**	**.20**	**.15**	**.10**	**.05**	**.025**	**.02**	**.01**	**.005**	**.0025**	**.001**	**.0005**
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27	50.51
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73	52.00
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18	53.48
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62	54.95
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05	56.41
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48	57.86
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89	59.30
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30	60.73
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70	62.16
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40	76.09
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66	89.56
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61	102.7
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8	128.3
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4	153.2

جدول (م5): جدول توزيع F



		γ_1								
γ_2	α	1	2	3	4	5	6	7	8	9
1	.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
	.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28
	.010	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
	.001	405284	500000	540379	562500	576405	585937	592873	598144	602284
2	.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
	.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
	.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39
3	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86
4	.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47
5	.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
	.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24
6	.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
	.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
	.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69
7	.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
	.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
	.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
	.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33

تابع جدول (م5)

γ_1										
10	**12**	**15**	**20**	**25**	**30**	**40**	**50**	**60**	**120**	**1000**
60.19	**60.71**	**61.22**	**61.74**	**62.05**	**62.26**	**62.53**	**62.69**	**62.79**	**63.06**	**63.30**
241.88	**243.91**	**245.95**	**248.01**	**249.26**	**250.10**	**251.14**	**251.77**	**252.20**	**253.25**	**254.19**
968.63	**976.71**	**984.87**	**993.10**	**998.08**	**1001.4**	**1005.6**	**1008.1**	**1009.8**	**1014.0**	**1017.7**
6055.8	**6106.3**	**6157.3**	**6208.7**	**6239.8**	**6260.6**	**6286.8**	**6302.5**	**6313.0**	**6339.4**	**6362.7**
605621	**610668**	**615764**	**620908**	**624017**	**626099**	**628712**	**630285**	**631337**	**633972**	**636301**
9.39	**9.41**	**9.42**	**9.44**	**9.45**	**9.46**	**9.47**	**9.47**	**9.47**	**9.48**	**9.49**
19.40	**19.41**	**19.43**	**19.45**	**19.46**	**19.46**	**19.47**	**19.48**	**19.48**	**19.49**	**19.49**
39.40	**39.41**	**39.43**	**39.45**	**39.46**	**39.46**	**39.47**	**39.48**	**39.48**	**39.49**	**39.50**
99.40	**99.42**	**99.43**	**99.45**	**99.46**	**99.47**	**99.47**	**99.48**	**99.48**	**99.49**	**99.50**
999.40	**999.42**	**999.43**	**999.45**	**999.46**	**999.47**	**999.47**	**999.48**	**999.48**	**999.49**	**999.50**
5.23	**5.22**	**5.20**	**5.18**	**5.17**	**5.17**	**5.16**	**5.15**	**5.15**	**5.14**	**5.13**
8.79	**8.74**	**8.70**	**8.66**	**8.63**	**8.62**	**8.59**	**8.58**	**8.57**	**8.55**	**8.53**
14.42	**14.34**	**14.25**	**14.17**	**14.12**	**14.08**	**14.04**	**14.01**	**13.99**	**13.95**	**13.91**
27.23	**27.05**	**26.87**	**26.69**	**26.58**	**26.50**	**26.41**	**26.35**	**26.32**	**26.22**	**26.14**
129.25	**128.32**	**127.37**	**126.42**	**125.84**	**125.45**	**124.96**	**124.66**	**124.47**	**123.97**	**123.53**
3.92	**3.90**	**3.87**	**3.84**	**3.83**	**3.82**	**3.80**	**3.80**	**3.79**	**3.78**	**3.76**
5.96	**5.91**	**5.86**	**5.80**	**5.77**	**5.75**	**5.72**	**5.70**	**5.69**	**5.66**	**5.63**
8.84	**8.75**	**8.66**	**8.56**	**8.50**	**8.46**	**8.41**	**8.38**	**8.36**	**8.31**	**8.26**
14.55	**14.37**	**14.20**	**14.02**	**13.91**	**13.84**	**13.75**	**13.69**	**13.65**	**13.56**	**13.47**
48.05	**47.41**	**46.76**	**46.10**	**45.70**	**45.43**	**45.09**	**44.88**	**44.75**	**44.40**	**44.09**
3.30	**3.27**	**3.24**	**3.21**	**3.19**	**3.17**	**3.16**	**3.15**	**3.14**	**3.12**	**3.11**
4.74	**4.68**	**4.62**	**4.56**	**4.52**	**4.50**	**4.46**	**4.44**	**4.43**	**4.40**	**4.37**
6.62	**6.52**	**6.43**	**6.33**	**6.27**	**6.23**	**6.18**	**6.14**	**6.12**	**6.07**	**6.02**
10.05	**9.89**	**9.72**	**9.55**	**9.45**	**9.38**	**9.29**	**9.24**	**9.20**	**9.11**	**9.03**
26.92	**26.42**	**25.91**	**25.39**	**25.08**	**24.87**	**24.60**	**24.44**	**24.33**	**24.06**	**23.82**
2.94	**2.90**	**2.87**	**2.84**	**2.81**	**2.80**	**2.78**	**2.77**	**2.76**	**2.74**	**2.72**
4.06	**4.00**	**3.94**	**3.87**	**3.83**	**3.81**	**3.77**	**3.75**	**3.74**	**3.70**	**3.67**
5.46	**5.37**	**5.27**	**5.17**	**5.11**	**5.07**	**5.01**	**4.98**	**4.96**	**4.90**	**4.86**
7.87	**7.72**	**7.56**	**7.40**	**7.30**	**7.23**	**7.14**	**7.09**	**7.06**	**6.97**	**6.89**
18.41	**17.99**	**17.56**	**17.12**	**16.85**	**16.67**	**16.44**	**16.31**	**16.21**	**15.98**	**15.77**
2.70	**2.67**	**2.63**	**2.59**	**2.57**	**2.56**	**2.54**	**2.52**	**2.51**	**2.49**	**2.47**
3.64	**3.57**	**3.51**	**3.44**	**3.40**	**3.38**	**3.34**	**3.32**	**3.30**	**3.27**	**3.23**
4.76	**4.67**	**4.57**	**4.47**	**4.40**	**4.36**	**4.31**	**4.28**	**4.25**	**4.20**	**4.15**
6.62	**6.47**	**6.31**	**6.16**	**6.06**	**5.99**	**5.91**	**5.86**	**5.82**	**5.74**	**5.66**
14.08	**13.71**	**13.32**	**12.93**	**12.69**	**12.53**	**12.33**	**12.20**	**12.12**	**11.91**	**11.72**

تابع جدول (م5)

			γ_1								
		α	1	2	3	4	5	6	7	8	9
γ_2	8	.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
		.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
		.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
		.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
		.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.77
	9	.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
		.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
		.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
		.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
		.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11
	10	.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
		.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
		.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
		.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
		.001	21.04	14.91	12.55	11.28	10.48	9.93	9.52	9.20	8.96
	11	.100	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
		.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
		.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
		.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
		.001	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12
	12	.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
		.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
		.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
		.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
		.001	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48
	13	.100	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
		.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
		.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
		.010	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
		.001	17.82	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98
	14	.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
		.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
		.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
		.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
		.001	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.58
	15	.100	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
		.050	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
		.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
		.010	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
		.001	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26
	16	.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
		.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
		.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
		.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
		.001	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.19	5.98
	17	.100	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
		.050	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
		.025	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
		.010	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68
		.001	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75

تابع جدول (م5)

γ_1										
10	**12**	**15**	**20**	**25**	**30**	**40**	**50**	**60**	**120**	**1000**
2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.35	2.34	2.32	2.30
3.35	3.28	3.22	3.15	3.11	3.08	3.04	3.02	3.01	2.97	2.93
4.30	4.20	4.10	4.00	3.94	3.89	3.84	3.81	3.78	3.73	3.68
5.81	5.67	5.52	5.36	5.26	5.20	5.12	5.07	5.03	4.95	4.87
11.54	11.19	10.84	10.48	10.26	10.11	9.92	9.80	9.73	9.53	9.36
2.42	2.38	2.34	2.30	2.27	2.25	2.23	2.22	2.21	2.18	2.16
3.14	3.07	3.01	2.94	2.89	2.86	2.83	2.80	2.79	2.75	2.71
3.96	3.87	3.77	3.67	3.60	3.56	3.51	3.47	3.45	3.39	3.34
5.26	5.11	4.96	4.81	4.71	4.65	4.57	4.52	4.48	4.40	4.32
9.89	9.57	9.24	8.90	8.69	8.55	8.37	8.26	8.19	8.00	7.84
2.32	2.28	2.24	2.20	2.17	2.16	2.13	2.12	2.11	2.08	2.06
2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.64	2.62	2.58	2.54
3.72	3.62	3.52	3.42	3.35	3.31	3.26	3.22	3.20	3.14	3.09
4.85	4.71	4.56	4.41	4.31	4.25	4.17	4.12	4.08	4.00	3.92
8.75	8.45	8.13	7.80	7.60	7.47	7.30	7.19	7.12	6.94	6.78
2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.04	2.03	2.00	1.98
2.85	2.79	2.72	2.65	2.60	2.57	2.53	2.51	2.49	2.45	2.41
3.53	3.43	3.33	3.23	3.16	3.12	3.06	3.03	3.00	2.94	2.89
4.54	4.40	4.25	4.10	4.01	3.94	3.86	3.81	3.78	3.69	3.61
7.92	7.63	7.32	7.01	6.81	6.68	6.52	6.42	6.35	6.18	6.02
2.19	2.15	2.10	2.06	2.03	2.01	1.99	1.97	1.96	1.93	1.91
2.75	2.69	2.62	2.54	2.50	2.47	2.43	2.40	2.38	2.34	2.30
3.37	3.28	3.18	3.07	3.01	2.96	2.91	2.87	2.85	2.79	2.73
4.30	4.16	4.01	3.86	3.76	3.70	3.62	3.57	3.54	3.45	3.37
7.29	7.00	6.71	6.40	6.22	6.09	5.93	5.83	5.76	5.59	5.44
2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.92	1.90	1.88	1.85
2.67	2.60	2.53	2.46	2.41	2.38	2.34	2.31	2.30	2.25	2.21
3.25	3.15	3.05	2.95	2.88	2.84	2.78	2.74	2.72	2.66	2.60
4.10	3.96	3.82	3.66	3.57	3.51	3.43	3.38	3.34	3.25	3.18
6.80	6.52	6.23	5.93	5.75	5.63	5.47	5.37	5.30	5.14	4.99
2.10	2.05	2.01	1.96	1.93	1.91	1.89	1.87	1.86	1.83	1.80
2.60	2.53	2.46	2.39	2.34	2.31	2.27	2.24	2.22	2.18	2.14
3.15	3.05	2.95	2.84	2.78	2.73	2.67	2.64	2.61	2.55	2.50
3.94	3.80	3.66	3.51	3.41	3.35	3.27	3.22	3.18	3.09	3.02
6.40	6.13	5.85	5.56	5.38	5.25	5.10	5.00	4.94	4.77	4.62
2.06	2.02	1.97	1.92	1.89	1.87	1.85	1.83	1.82	1.79	1.76
2.54	2.48	2.40	2.33	2.28	2.25	2.20	2.18	2.16	2.11	2.07
3.06	2.96	2.86	2.76	2.69	2.64	2.59	2.55	2.52	2.46	2.40
3.80	3.67	3.52	3.37	3.28	3.21	3.13	3.08	3.05	2.96	2.88
6.08	5.81	5.54	5.25	5.07	4.95	4.80	4.70	4.64	4.47	4.33
2.03	1.99	1.94	1.89	1.86	1.84	1.81	1.79	1.78	1.75	1.72
2.49	2.42	2.35	2.28	2.23	2.19	2.15	2.12	2.11	2.06	2.02
2.99	2.89	2.79	2.68	2.61	2.57	2.51	2.47	2.45	2.38	2.32
3.69	3.55	3.41	3.26	3.16	3.10	3.02	2.97	2.93	2.84	2.76
5.81	5.55	5.27	4.99	4.82	4.70	4.54	4.45	4.39	4.23	4.08
2.00	1.96	1.91	1.86	1.83	1.81	1.78	1.76	1.75	1.72	1.69
2.45	2.38	2.31	2.23	2.18	2.15	2.10	2.08	2.06	2.01	1.97
2.92	2.82	2.72	2.62	2.55	2.50	2.44	2.41	2.38	2.32	2.26
3.59	3.46	3.31	3.16	3.07	3.00	2.92	2.87	2.83	2.75	2.66
5.58	5.32	5.05	4.78	4.60	4.48	4.33	4.24	4.18	4.02	3.87

تابع جدول (م5)

γ_2	α	γ_1 = 1	2	3	4	5	6	7	8	9
	.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
	.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
18	.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
	.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
	.001	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56
	.100	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
	.050	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
19	.025	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
	.010	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
	.001	15.08	10.16	8.28	7.27	6.62	6.18	5.85	5.59	5.39
	.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
	.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
20	.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
	.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
	.001	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24
	.100	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
	.050	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
21	.025	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
	.010	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
	.001	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11
	.100	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
	.050	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
22	.025	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
	.010	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
	.001	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99
	.100	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
	.050	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
23	.025	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
	.010	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
	.001	14.20	9.47	7.67	6.70	6.08	5.65	5.33	5.09	4.89
	.100	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
	.050	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
24	.025	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
	.010	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
	.001	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80
	.100	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
	.050	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
25	.025	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
	.010	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
	.001	13.88	9.22	7.45	6.49	5.89	5.46	5.15	4.91	4.71
	.100	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
	.050	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
26	.025	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
	.010	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
	.001	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64
	.100	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
	.050	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
27	.025	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
	.010	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
	.001	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57

تابع جدول (م5)

γ_1										
10	**12**	**15**	**20**	**25**	**30**	**40**	**50**	**60**	**120**	**1000**
1.98	1.93	1.89	1.84	1.80	1.78	1.75	1.74	1.72	1.69	1.66
2.41	2.34	2.27	2.19	2.14	2.11	2.06	2.04	2.02	1.97	1.92
2.87	2.77	2.67	2.56	2.49	2.44	2.38	2.35	2.32	2.26	2.20
3.51	3.37	3.23	3.08	2.98	2.92	2.84	2.78	2.75	2.66	2.58
5.39	5.13	4.87	4.59	4.42	4.30	4.15	4.06	4.00	3.84	3.69
1.96	1.91	1.86	1.81	1.78	1.76	1.73	1.71	1.70	1.67	1.64
2.38	2.31	2.23	2.16	2.11	2.07	2.03	2.00	1.98	1.93	1.88
2.82	2.72	2.62	2.51	2.44	2.39	2.33	2.30	2.27	2.20	2.14
3.43	3.30	3.15	3.00	2.91	2.84	2.76	2.71	2.67	2.58	2.50
5.22	4.97	4.70	4.43	4.26	4.14	3.99	3.90	3.84	3.68	3.53
1.94	1.89	1.84	1.79	1.76	1.74	1.71	1.69	1.68	1.64	1.61
2.35	2.28	2.20	2.12	2.07	2.04	1.99	1.97	1.95	1.90	1.85
2.77	2.68	2.57	2.46	2.40	2.35	2.29	2.25	2.22	2.16	2.09
3.37	3.23	3.09	2.94	2.84	2.78	2.69	2.64	2.61	2.52	2.43
5.08	4.82	4.56	4.29	4.12	4.00	3.86	3.77	3.70	3.54	3.40
1.92	1.87	1.83	1.78	1.74	1.72	1.69	1.67	1.66	1.62	1.59
2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.94	1.92	1.87	1.82
2.73	2.64	2.53	2.42	2.36	2.31	2.25	2.21	2.18	2.11	2.05
3.31	3.17	3.03	2.88	2.79	2.72	2.64	2.58	2.55	2.46	2.37
4.95	4.70	4.44	4.17	4.00	3.88	3.74	3.64	3.58	3.42	3.28
1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.65	1.64	1.60	1.57
2.30	2.23	2.15	2.07	2.02	1.98	1.94	1.91	1.89	1.84	1.79
2.70	2.60	2.50	2.39	2.32	2.27	2.21	2.17	2.14	2.08	2.01
3.26	3.12	2.98	2.83	2.73	2.67	2.58	2.53	2.50	2.40	2.32
4.83	4.58	4.33	4.06	3.89	3.78	3.63	3.54	3.48	3.32	3.17
1.89	1.84	1.80	1.74	1.71	1.69	1.66	1.64	1.62	1.59	1.55
2.27	2.20	2.13	2.05	2.00	1.96	1.91	1.88	1.86	1.81	1.76
2.67	2.57	2.47	2.36	2.29	2.24	2.18	2.14	2.11	2.04	1.98
3.21	3.07	2.93	2.78	2.69	2.62	2.54	2.48	2.45	2.35	2.27
4.73	4.48	4.23	3.96	3.79	3.68	3.53	3.44	3.38	3.22	3.08
1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.62	1.61	1.57	1.54
2.25	2.18	2.11	2.03	1.97	1.94	1.89	1.86	1.84	1.79	1.74
2.64	2.54	2.44	2.33	2.26	2.21	2.15	2.11	2.08	2.01	1.94
3.17	3.03	2.89	2.74	2.64	2.58	2.49	2.44	2.40	2.31	2.22
4.64	4.39	4.14	3.87	3.71	3.59	3.45	3.36	3.29	3.14	2.99
1.87	1.82	1.77	1.72	1.68	1.66	1.63	1.61	1.59	1.56	1.52
2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.84	1.82	1.77	1.72
2.61	2.51	2.41	2.30	2.23	2.18	2.12	2.08	2.05	1.98	1.91
3.13	2.99	2.85	2.70	2.60	2.54	2.45	2.40	2.36	2.27	2.18
4.56	4.31	4.06	3.79	3.63	3.52	3.37	3.28	3.22	3.06	2.91
1.86	1.81	1.76	1.71	1.67	1.65	1.61	1.59	1.58	1.54	1.51
2.22	2.15	2.07	1.99	1.94	1.90	1.85	1.82	1.80	1.75	1.70
2.59	2.49	2.39	2.28	2.21	2.16	2.09	2.05	2.03	1.95	1.89
3.09	2.96	2.81	2.66	2.57	2.50	2.42	2.36	2.33	2.23	2.14
4.48	4.24	3.99	3.72	3.56	3.44	3.30	3.21	3.15	2.99	2.84
1.85	1.80	1.75	1.70	1.66	1.64	1.60	1.58	1.57	1.53	1.50
2.20	2.13	2.06	1.97	1.92	1.88	1.84	1.81	1.79	1.73	1.68
2.57	2.47	2.36	2.25	2.18	2.13	2.07	2.03	2.00	1.93	1.86
3.06	2.93	2.78	2.63	2.54	2.47	2.38	2.33	2.29	2.20	2.11
4.41	4.17	3.92	3.66	3.49	3.38	3.23	3.14	3.08	2.92	2.78

تابع جدول (م5)

γ_2	α	γ_1 = 1	2	3	4	5	6	7	8	9
28	.100	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
	.050	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
	.025	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
	.010	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
	.001	13.50	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50
29	.100	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
	.050	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
	.025	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
	.010	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
	.001	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45
30	.100	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
	.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
	.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
	.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
	.001	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39
40	.100	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
	.050	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
	.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
	.010	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
	.001	12.61	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02
50	.100	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76
	.050	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07
	.025	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.38
	.010	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78
	.001	12.22	7.96	6.34	5.46	4.90	4.51	4.22	4.00	3.82
60	.100	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
	.050	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
	.025	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
	.010	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
	.001	11.97	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69
100	.100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69
	.050	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97
	.025	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.24
	.010	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59
	.001	11.50	7.41	5.86	5.02	4.48	4.11	3.83	3.61	3.44
200	.100	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66
	.050	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93
	.025	5.10	3.76	3.18	2.85	2.63	2.47	2.35	2.26	2.18
	.010	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50
	.001	11.15	7.15	5.63	4.81	4.29	3.92	3.65	3.43	3.26
1000	.100	2.71	2.31	2.09	1.95	1.85	1.78	1.72	1.68	1.64
	.050	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89
	.025	5.04	3.70	3.13	2.80	2.58	2.42	2.30	2.20	2.13
	.010	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43
	.001	10.89	6.96	5.46	4.65	4.14	3.78	3.51	3.30	3.13

تابع جدول (م5)

γ_1										
10	**12**	**15**	**20**	**25**	**30**	**40**	**50**	**60**	**120**	**1000**
1.84	1.79	1.74	1.69	1.65	1.63	1.59	1.57	1.56	1.52	1.48
2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.79	1.77	1.71	1.66
2.55	2.45	2.34	2.23	2.16	2.11	2.05	2.01	1.98	1.91	1.84
3.03	2.90	2.75	2.60	2.51	2.44	2.35	2.30	2.26	2.17	2.08
4.35	4.11	3.86	3.60	3.43	3.32	3.18	3.09	3.02	2.86	2.72
1.83	1.78	1.73	1.68	1.64	1.62	1.58	1.56	1.55	1.51	1.47
2.18	2.10	2.03	1.94	1.89	1.85	1.81	1.77	1.75	1.70	1.65
2.53	2.43	2.32	2.21	2.14	2.09	2.03	1.99	1.96	1.89	1.82
3.00	2.87	2.73	2.57	2.48	2.41	2.33	2.27	2.23	2.14	2.05
4.29	4.05	3.80	3.54	3.38	3.27	3.12	3.03	2.97	2.81	2.66
1.82	1.77	1.72	1.67	1.63	1.61	1.57	1.55	1.54	1.50	1.46
2.16	2.09	2.01	1.93	1.88	1.84	1.79	1.76	1.74	1.68	1.63
2.51	2.41	2.31	2.20	2.12	2.07	2.01	1.97	1.94	1.87	1.80
2.98	2.84	2.70	2.55	2.45	2.39	2.30	2.25	2.21	2.11	2.02
4.24	4.00	3.75	3.49	3.33	3.22	3.07	2.98	2.92	2.76	2.61
1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.48	1.47	1.42	1.38
2.08	2.00	1.92	1.84	1.78	1.74	1.69	1.66	1.64	1.58	1.52
2.39	2.29	2.18	2.07	1.99	1.94	1.88	1.83	1.80	1.72	1.65
2.80	2.66	2.52	2.37	2.27	2.20	2.11	2.06	2.02	1.92	1.82
3.87	3.64	3.40	3.14	2.98	2.87	2.73	2.64	2.57	2.41	2.25
1.73	1.68	1.63	1.57	1.53	1.50	1.46	1.44	1.42	1.38	1.33
2.03	1.95	1.87	1.78	1.73	1.69	1.63	1.60	1.58	1.51	1.45
2.32	2.22	2.11	1.99	1.92	1.87	1.80	1.75	1.72	1.64	1.56
2.70	2.56	2.42	2.27	2.17	2.10	2.01	1.95	1.91	1.80	1.70
3.67	3.44	3.20	2.95	2.79	2.68	2.53	2.44	2.38	2.21	2.05
1.71	1.66	1.60	1.54	1.50	1.48	1.44	1.41	1.40	1.35	1.30
1.99	1.92	1.84	1.75	1.69	1.65	1.59	1.56	1.53	1.47	1.40
2.27	2.17	2.06	1.94	1.87	1.82	1.74	1.70	1.67	1.58	1.49
2.63	2.50	2.35	2.20	2.10	2.03	1.94	1.88	1.84	1.73	1.62
3.54	3.32	3.08	2.83	2.67	2.55	2.41	2.32	2.25	2.08	1.92
1.66	1.61	1.56	1.49	1.45	1.42	1.38	1.35	1.34	1.28	1.22
1.93	1.85	1.77	1.68	1.62	1.57	1.52	1.48	1.45	1.38	1.30
2.18	2.08	1.97	1.85	1.77	1.71	1.64	1.59	1.56	1.46	1.36
2.50	2.37	2.22	2.07	1.97	1.89	1.80	1.74	1.69	1.57	1.45
3.30	3.07	2.84	2.59	2.43	2.32	2.17	2.08	2.01	1.83	1.64
1.63	1.58	1.52	1.46	1.41	1.38	1.34	1.31	1.29	1.23	1.16
1.88	1.80	1.72	1.62	1.56	1.52	1.46	1.41	1.39	1.30	1.21
2.11	2.01	1.90	1.78	1.70	1.64	1.56	1.51	1.47	1.37	1.25
2.41	2.27	2.13	1.97	1.87	1.79	1.69	1.63	1.58	1.45	1.30
3.12	2.90	2.67	2.42	2.26	2.15	2.00	1.90	1.83	1.64	1.43
1.61	1.55	1.49	1.43	1.38	1.35	1.30	1.27	1.25	1.18	1.08
1.84	1.76	1.68	1.58	1.52	1.47	1.41	1.36	1.33	1.24	1.11
2.06	1.96	1.85	1.72	1.64	1.58	1.50	1.45	1.41	1.29	1.13
2.34	2.20	2.06	1.90	1.79	1.72	1.61	1.54	1.50	1.35	1.16
2.99	2.77	2.54	2.30	2.14	2.02	1.87	1.77	1.69	1.49	1.22

ملحق 2

حلول تمارين الكتاب

1. حلول تمارين الفصل الأول

تمرين (1.1): المتغير في هذه الحالة هو عدد الأطفال في سن الدراسة تحت 18 سنة وهو متغير كمي مقاس بالسنة، وعدد المشاهدات هو 250 أسرة.

تمرين (2.1): المتغير الأول يمثل عدد الساعات المستغرقة في العمل الفعلي لموظفي الحسابات الجارية وهو متغير كمي مقاس بالساعة، والمتغير الثاني هو اسم الفرع الذي يعمل به الموظف، (والذي يمكن أن يتم التعبير عنه برمز أو رقم)، وهو متغير وصفي. وحجم العينة هو 150 موظف.

تمرين (3.1): الدراسة تصنف ضمن الإحصاء الاستدلالي لأن النتيجة المذكورة تم حسابها من عينة مسحوبة من مجتمع حالات الطلاق ولا تشمل كل حالات الطلاق (المجتمع).

تمرين (4.1): المتغيرات في (ب)، (ج)، و (د) هي متغيرات وصفية.

تمرين (5.1): (أ) مقياس اسمي، (ب) مقياس نسبي، (ج) مقياس رتبي أو ترتيبي، و (د) مقياس فئوي.

2. حلول تمارين الفصل الثاني

تمرين (1.2):

التكرار f	الفترة (ضريبة الدخل)
2	100 – 125
5	125 – 150
10	150 – 175
10	175 – 200
4	200 – 225
4	225 – 250

تمرين (2.2):

أ.

النسبة المئوية	التكرار النسبي	التكرار f	آراء المواطنين في أداء قطاع الصحة
16%	0.16	4	سيء
48%	0.48	12	مقبول
24%	0.24	6	جيد
12%	0.12	3	ممتاز

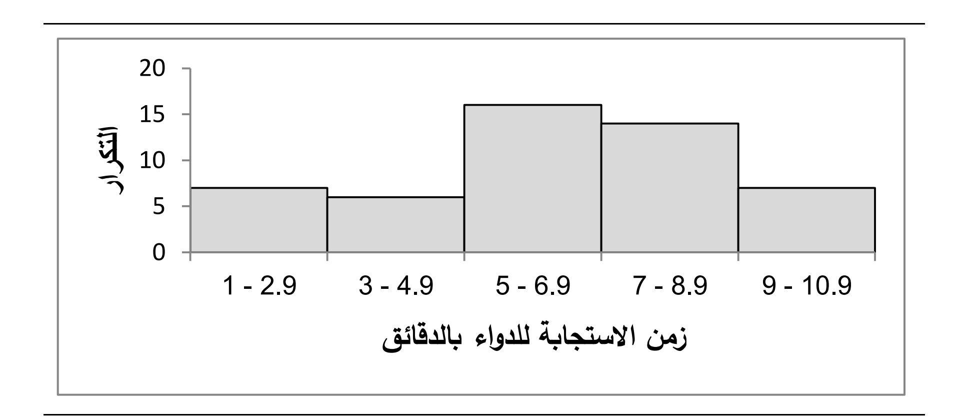
ب. 36% = 12% + 24%

تمرين (3.2):

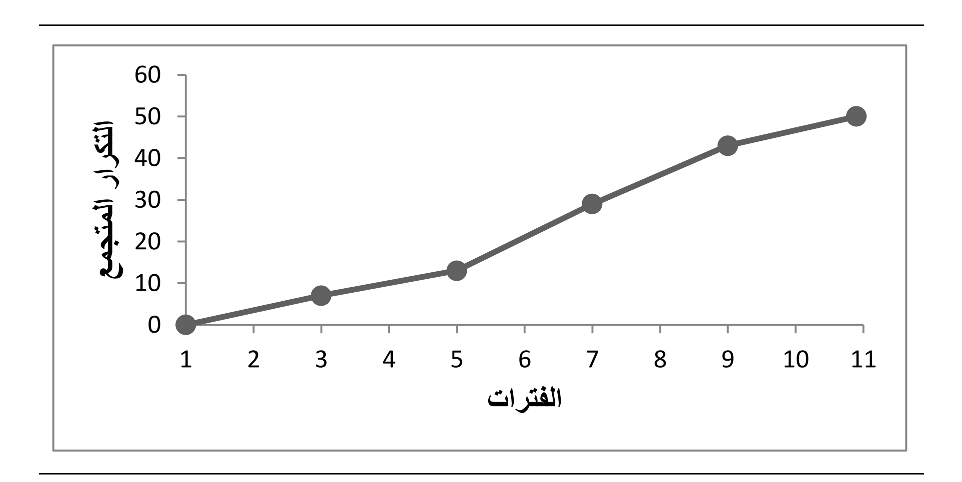
ب				أ	
التكرار المتجمع الصاعد					
النسبة التراكمية	النسبي	التكرار	الفترة	التكرار f	الفترة
0%	0	0	أقل من 1		
14%	0.14	7	أقل من 3	7	1 – 2.9
26%	0.26	13	أقل من 5	6	3 – 4.9
58%	0.58	29	أقل من 7	16	5 – 6.9
86%	0.86	43	أقل من 9	14	7 – 8.9
100%	1	50	أقل من 11	7	9 – 10.9

تمرين (4.2):

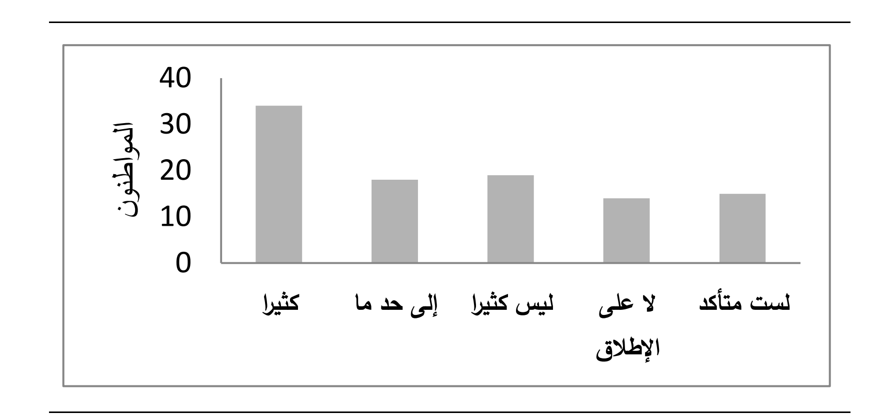
أ.

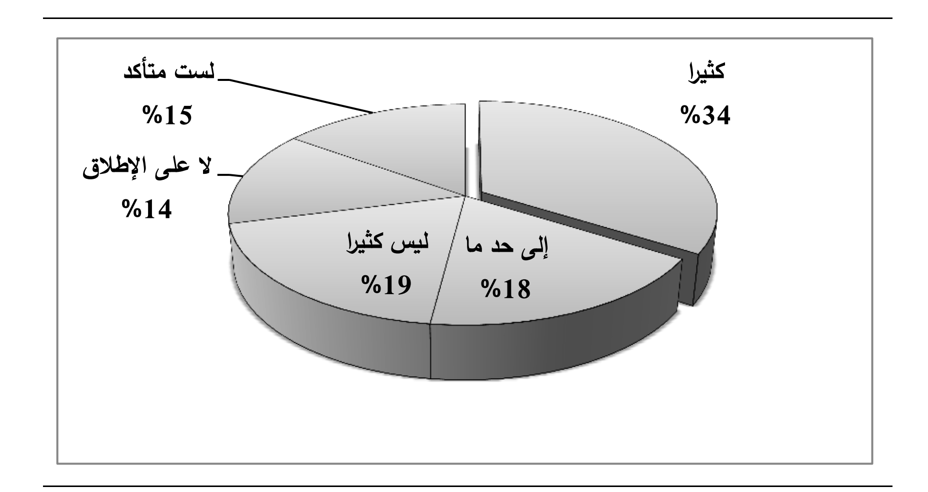


ب.



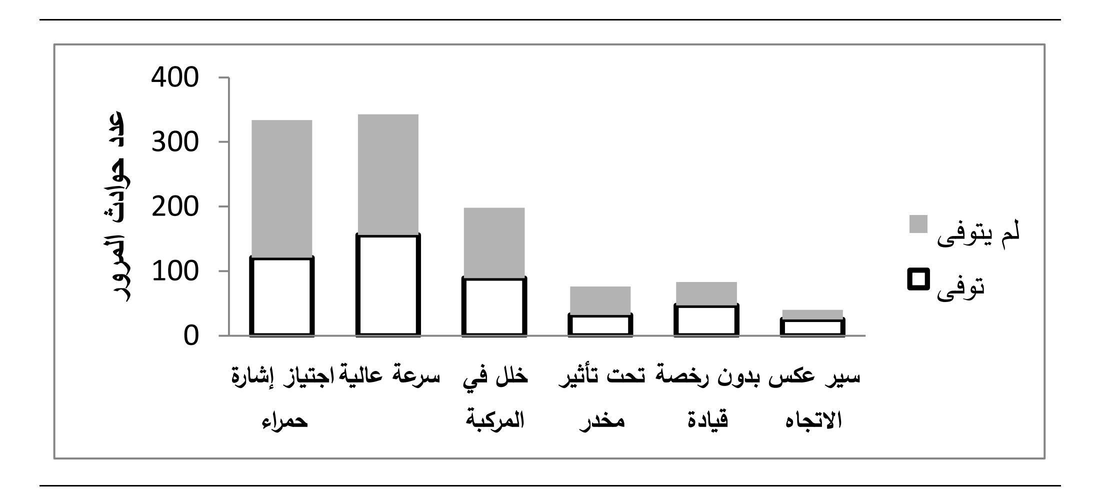
تمرين (5.2): الأعمدة البيانية والقطاعات الدائرية لإجابات المواطنين؛



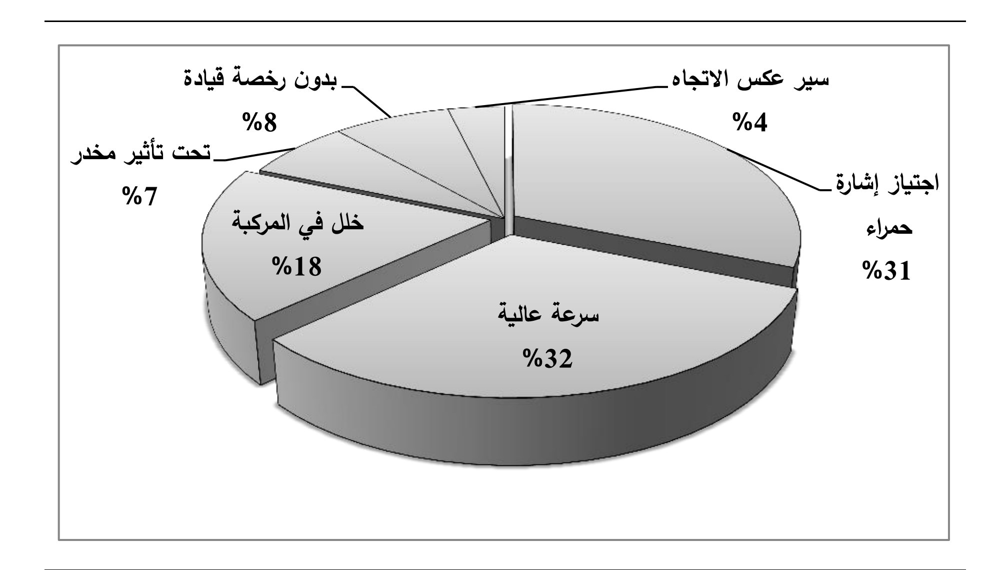


تمرين (6.2):

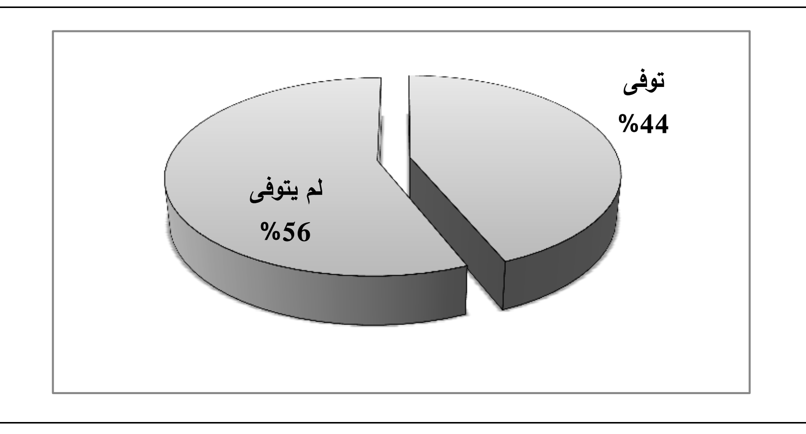
أ.



ب.



ج.



تمرين (7.2): أ. 0.71−   ب. 1−   ج. 1− .

تمرين (8.2): أ. 13.79   ب. 11.89   ج. 9.45   د. 13.79   هـ. 12.67
و. 13.37   ز. 18.36   ح. 10.10   ط. 18.36.

3. حلول تمارين الفصل الثالث

تمرين (1.3):

أ.

1. المدى = 25−18 = 7، 2. المدى الربيعي = 23.5−20 = 3.5، 3. نصف المدى الربيعي=3.5/2=1.75، 4. الانحراف المتوسط = 1.9، 5. الانحراف المعياري = 2.20، التباين = 4.85 .

ب.

C.V(بنغازي) = (2.20/21.5)×100 = 10.24
C.V(الاسكندرية) = (3.96/70.7)×100 = 5.61

ملاحظة: قد يبدو من النظرة الأولى أن الاختلاف في درجات الحرارة في مدينة بنغازي هو أكبر مما هو عليه في الاسكندرية، إلا أن ذلك ليس صحيح لأن درجات الحرارة في جدول 2 هي نفس الدرجات في جدول 1 ولكنها مقاسة بالفهرنهايت، لذلك فإنه يجب في مثل هذه الحالات توحيد وحدة القياس (ويمكن ذلك باستخدام التحويل؛ (الدرجة بالفهرنهايت−32)×5/9)، أو استخدام الدرجات المعيارية.

تمرين (2.3): باستخدام الجدول التالي للحسابات:

الفترة المستمرة	التكرار f	مركز الفترة x	$f_i x_i$	$\lvert x_i - \bar{X}\rvert$	$f_i\lvert x_i - \bar{X}\rvert$	$f_i x_i^2$
5 – 15	25	10	250	16.63	415.83	2500
15 – 25	120	20	2400	6.63	796.00	48000
25 – 35	105	30	3150	3.37	353.50	94500
35 – 45	31	40	1240	13.37	414.37	49600
45 – 55	19	50	950	23.37	443.97	47500
المجموع	300		7990		2423.67	242100

1. المدى = 55 − 5 = 50، 2. نصف المدى الربيعي = 2/(32.62 − 19.17) = 6.73، 3. الانحراف المتوسط = 2423.67/300 = 8.08، 4. الانحراف المعياري = 9.88 ، والتباين = 97.67 5. المدى المئيني = 45.65 − 25.04 = 20.60 .

تمرين (3.3):

بنغازي	-0.23	1.59	1.14	0.23	0.23	-1.59	-0.68	1.14	-1.14	-0.68
الاسكندرية	-0.23	1.59	1.14	0.23	0.23	-1.59	-0.68	1.14	-1.14	-0.68

درجات الحرارة في كلا المدينتين لها نفس درجة التشتت تماما، والاختلاف في البيانات الأصلية كان نتيجة التحويل من درجة مئوية إلى فهرنهايت. ويكون

$$\bar{X}_1 = \bar{X}_2 = 0 \text{ ، } SD_1 = SD_2 = 1 \text{ .}$$

تمرين (4.3): لدينا من الشركة A و الشركة B؛ $Z_A = 0.31$, $Z_B = 1.19$ ، وبالتالي راتب الموظف في الشركة B (850 دينار) هو أفضل مقارنة بمستوى المرتبات في الشركة.

تمرين (5.3):

أ.

$$\acute{\mu}_1 = 3.9 \text{ , } \acute{\mu}_2 = 106.9 \text{ , } \acute{\mu}_3 = 812.7 \text{ , } \acute{\mu}_4 = 29350.9$$

ب.

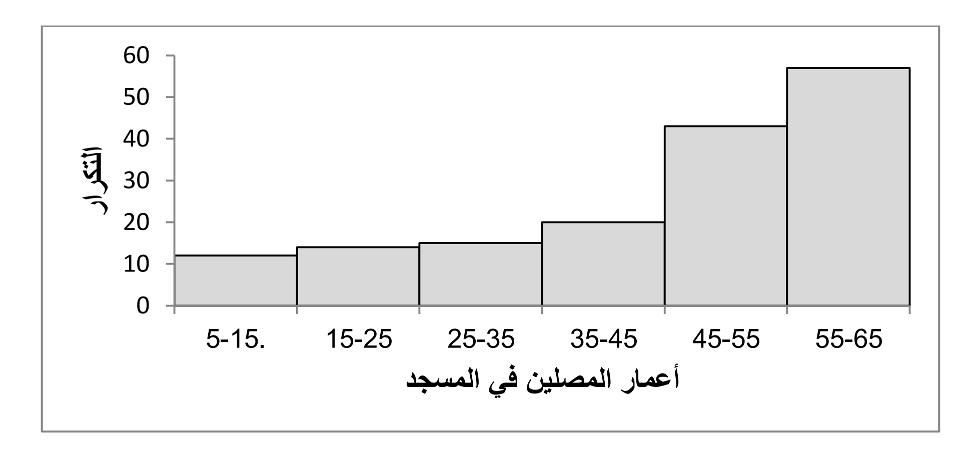
$$\bar{X} = \acute{\mu}_1 = 3.9 \text{ , } Var(X) = \acute{\mu}_2 - \acute{\mu}_1^2 = 91.69$$

ج.

$$\mu_1 = 0 \text{ , } \mu_2 = 91.69 \text{ , } \mu_3 = -319.39 \text{ , } \mu_4 = 25734.44$$

تمرين (6.3):

باستخدام معامل الالتواء العزمي؛ $S.K_{\mu} = \frac{(-356.58)^2}{(16.01)^3} = 31.04$ ، وبالنظر إلى إشارة العزم الثالث حول الوسط الحسابي في البسط نجدها سالبة، مما يدل على وجود التواء حاد إلى اليسار وهذا ما يظهره شكل المدرج التكراري للبيانات.



وبحساب معامل التفرطح العزمي؛ $Kur_{\mu} = \frac{12160}{[16.01]^4} = 0.19$ ، مما يدل على وجود تفرطح واضح في البيانات.

تمرين (7.3):

من الشكل يتضح أن توزيع أعداد المشتركين في الشركة 1 والشركة 2 هي تقريبا متساوية، أما الشركة 3 فإن أعداد المشتركين فيها تزيد بشكل ملحوظ عن باقي الشركات مع وجود تشتت أكبر في توزيع أعداد المشتركين فيها.

تمرين (8.3): من شكل الساق والورقة يُلاحظ أن توزيع درجات الذكاء يقترب كثيرا من التوزيع الطبيعي.

الساق	الورقة
2	0
3	0 3
4	0 4 5
5	0 2 3 5
6	0 1 2 5 5 6
7	0 2 4 8
8	4 8 9
9	1 2

4. حلول تمارين الفصل الرابع

تمرين (1.4): عدد النتائج الممكنة هو $5^3 = 125$ نتيجة كلية تأخذ النمط S = {AAA, AAB, AAC, AAD, …, FFF} .

تمرين (2.4): إذا ما تم استخدام الرموز التالية للأحداث؛ قسم الرياضيات = Ma ، قسم الإحصاء = St ، وقسم الكيمياء = Ch ، فإنه يمكن تكوين العلاقات التالية (على سبيل المثال لا الحصر):

Ma ∪ St = { رياضة2، رياضة1، إحصاء1، إحصاء2، لغة إنجليزية }

Ma ∩ Ch = { لغة إنجليزية }

St^c = { رياضة2، كيمياء1، نبات1، نبات2، فيزياء1، فيزياء2 }

$(Ma \cup St)^c$ = { كيمياء1، نبات1، نبات2، فيزياء1، فيزياء2 }

تمرين (3.4): يمكن تكوين $2^6 = 64$ تركيبة.

تمرين (4.4):

أ. $10^3 = 1000$ طريقة، ب. $P_3^{10} = 720$ طريقة، ج. $C_3^{10} = 120$ طريقة .

تمرين (5.4): $C_2^7 = 21$ طريقة على اعتبار أن لون السيارة غير مهم.

تمرين (6.4): $\frac{11!}{4!3!4!} = 11550$ طريقة .

تمرين (7.4):

بالنسبة للحساء فإنه يمكن اختيار 2 من 4 أنواع بعدد $C_2^4 = 6$ طرق ،

وبالنسبة للمقبلات فإنه يمكن اختيار 3 من بين 6 أنواع بعدد؛ $C_3^6 = 20$ طريقة ،

وبالنسبة للحوم فإنه يمكن اختيار 3 من بين 5 أنواع بعدد؛ $C_3^5 = 10$ طرق ،

وبالتالي فإن عدد الطرق الكلية للاختيار هو $6 \times 20 \times 10 = 1200$ طريقة.

تمرين (8.4):

أ. $5/25$ ب. $20/25$.

تمرين (9.4):

أ. $\frac{C_1^5 C_1^6}{C_2^{11}} = 0.55$ ب. $\frac{C_2^6}{C_2^{11}} = 0.27$ ج. $\frac{C_2^5}{C_2^{11}} + \frac{C_2^6}{C_2^{11}} = 0.45$.

تمرين (10.4): لتبسيط الحل يمكن تكوين الجدول التالي:

		نظام المكيف		
		نظام القطعة الواحدة	نظام القطعتين	المجموع
قوة المكيف	القوة 24	6	3	9
	القوة 12	4	7	11
	المجموع	10	10	20

أ. بوضع A هو حدث أن كلا المكيفين من نظام القطعة الواحدة فيكون

$P(A) = \frac{C_2^{10}}{C_2^{20}} = 0.24$.

ب. بوضع B هو حدث أن كلا المكيفين قوته 24 فيكون $P(B) = \frac{C_2^9}{C_2^{20}} = 0.19$.

ج. $P(A \cap B) = \frac{C_2^6}{C_2^{20}} = 0.08$.

د. $P(A \cup B) = 0.24 + 0.19 - 0.08 = 0.35$.

تمرين (11.4):

أ. $\frac{C_2^{10} \times C_3^5}{C_5^{15}} = 0.15$، ب. $\frac{C_3^5 \times C_2^{10}}{C_5^{15}} = 0.15$، ج. $\frac{C_1^5 \times C_4^{10} + C_2^5 \times C_3^{10} + C_3^5 \times C_2^{10}}{C_5^{15}} = 0.90$،

د. $\frac{C_2^5 \times C_3^{10} + C_1^5 \times C_4^{10} + C_0^5 \times C_5^{10}}{C_5^{15}} = 0.83$.

تمرين (12.4):

أ. $\frac{7/20}{10/20} = 0.70$ ، ب. $\frac{6/20}{9/20} = 0.67$ ، ج. $\frac{11}{20} + \frac{10}{20} - \frac{7}{20} = 0.70$.

تمرين (13.4):

أ. S = {A1A2A3, A1A2B3, A1B2A3, A1B2B3, B1A2A3, B1A2B3, {B1B2A3, B1B2B3، حيث A = ذكر ، B = أنثى ، و 1، 2، 3 ترمز للإنجاب الأول، الثاني، والثالث على التوالي.

ب. P(A2\B1) = P(A2∩B1)/P(B1) = P(A2) = 4/8 = 0.5 لأنها أحداث مستقلة.

تمرين (14.4):

أ. $\frac{9}{14}\times\frac{8}{13}\times\frac{7}{12}=0.23$ ب. $\frac{9}{14}\times\frac{5}{13}\times\frac{4}{12}=0.08$.

تمرين (15.4):

أ. 0.20 ب. 0.42 .

5. حلول تمارين الفصل الخامس

تمرين (1.5):

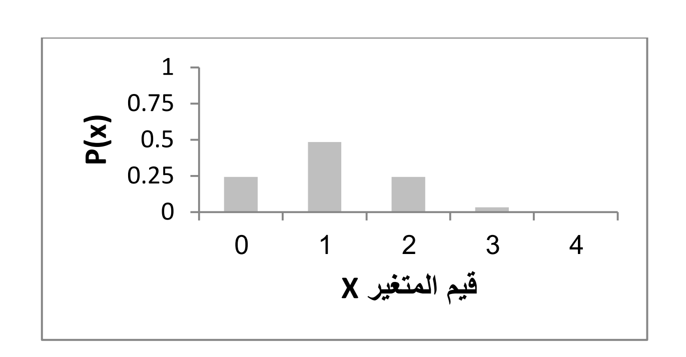
.1

x	0	1	2	3	4
$P(x)$	0.242	0.484	0.242	0.032	0.001

2. 0.242 ، 3. 0.242 + 0.484 = 0.726 ، 4. 0.032 + 0.001 = 0.033 ،

5. 0.242 + 0.032 + 0.001 = 0.275 ، 6. 0.242 + 0.484 + 0.242 = 0.968 ،

.7



تمرين (2.5):

1. $\int_0^3 \frac{1}{9}y^2dy=\frac{1}{9}\left[\frac{y^3}{3}\right]_0^3=\frac{1}{27}[27-0]=1$.

2. $P(Y>2)=P(2<Y\le 3)=\int_2^3 \frac{y^2}{9}dy=\frac{1}{9}\left[\frac{y^3}{3}\right]_2^3=\frac{1}{27}[27-8]=\frac{19}{27}$.

3. $P(0<Y\le 2)=\int_0^2 \frac{y^2}{9}dy=\frac{1}{9}\left[\frac{y^3}{3}\right]_0^2=\frac{1}{27}[8-0]=\frac{8}{27}$.

4. $P(2\le Y\le 5)=P(2\le Y\le 3)+P(3<Y\le 5)=\frac{19}{27}+0=\frac{19}{27}$.

5. $F(1)=P(Y\le 1)=\int_0^1 \frac{1}{9}y^2dy=\frac{1}{9}\left[\frac{y^3}{3}\right]_0^1=\frac{1}{27}[1-0]=\frac{1}{27}$ ،

$F(3)=F(max(y))=\int_0^3 \frac{1}{9}y^2dy=1$.

تمرين (3.5):

$$\int_0^3 (ax^2 + \frac{1}{9}x)dx = 1$$

$$a\left[\frac{x^3}{3}\right]_0^3 + \frac{1}{9}\left[\frac{x^2}{2}\right]_0^3 = 1$$

$$a\left[\frac{27}{3} - 0\right] + \frac{1}{9}\left[\frac{9}{2} - 0\right] = 1$$

$$9a + \frac{1}{2} = 1$$

$$9a = \frac{1}{2}$$

$$a = \frac{1}{18}$$

تمرين (4.5):

1. $\sum(\frac{1}{4} + \frac{1}{4} + \frac{1}{3} + k) = 1$ ، وبالتالي $k = \frac{1}{6}$.

2. $E(X) = 1.42$ 3. $E(Z) = 2.33$ 4. $Var(X) = 1.08$ 5. $Var(Y) = 0.12$

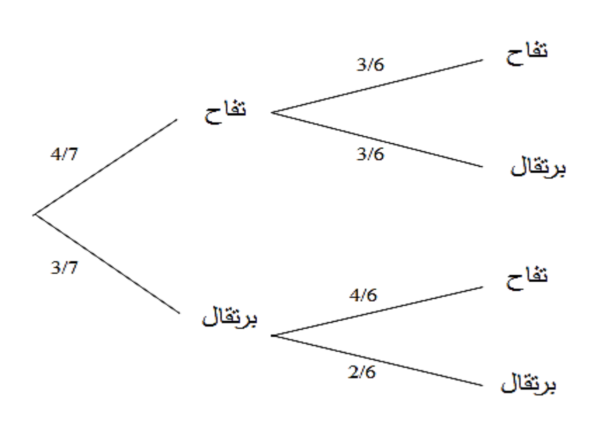
تمرين (5.5):

1. $E(X) = 2.125$ 2. $E(3X) = 6.357$ 3. $Var(X) = 4.95 - 4.52 = 0.43$.

4. $Var(2X - 3) = 1.72$.

تمرين (6.5):

.1



X_1 \ X_2	0	1	المجموع
0	6/42	12/42	18/42
1	12/42	12/42	24/42
المجموع	18/42	24/42	1

.2

X_1	0	1
$P_1(x_1)$	18/42	24/42

X_2	0	1
$P_2(x_2)$	18/42	24/42

.3 $P(0,0)=\frac{6}{42}\neq P_1(0)\times P_2(0)=\frac{18}{42}\times\frac{18}{42}=\frac{9}{49}$ ، غير مستقلان.

.4 $E(X_1)=\frac{24}{42}$ ، $E(X_2)=\frac{24}{42}$ ، $E(X_1X_2)=\frac{12}{42}$.

تمرين (7.5):

.1 $E(X_1)=\frac{14}{20}$ ، $E(X_2)=\frac{5}{20}$ و $E(X_1X_2)=\frac{2}{20}$.

.2 $Cov(X_1,X_2)=\frac{2}{20}-\frac{14}{20}\times\frac{5}{20}=-\frac{3}{40}$.

$\rho_{X_1X_2}=\frac{-\frac{3}{40}}{\sqrt{\frac{21}{100}\times\frac{3}{16}}}=-0.38$ ، توجد علاقة عكسية ضعيفة بين المتغيرين.

تمرين(8.5): $k=3$.

تمرين (9.5):

.1 $\acute{\mu}_1=0,\ \acute{\mu}_2=4,\ \acute{\mu}_3=0,\ \acute{\mu}_4=32.8$

.2 $\mu_1=0,\ \ \mu_2=4,\ \ \mu_3=0,\ \ \mu_4=32.8$

.3 $SK_\mu=0,\ \ Kur_\mu=2.05$

.4

$$\mu_x(t)=E(e^{tX})=\sum_x e^{tX}P(x)$$
$$=e^{-3t}\times\frac{1}{5}+e^{-1t}\times\frac{1}{5}+e^{0t}\times\frac{1}{5}+e^{1t}\times\frac{1}{5}+e^{3t}\times\frac{1}{5}$$
$$=\frac{1}{5}(e^{-3t}+e^{-t}+e^{t}+e^{3t}+1)$$

.5

$$\acute{\mu}_1 = \frac{\partial \mu_x(t)}{\partial t}\bigg|_{t=0} = \frac{1}{5}(-3e^{-3t} - e^{-t} + e^{t} + 3e^{3t} + 0)\bigg|_{t=0} = \frac{1}{5}(-3 - 1 + 1 + 3) = 0$$

$$\acute{\mu}_2 = \frac{\partial^2 \mu_x(t)}{\partial t^2}\bigg|_{t=0} = \frac{1}{5}(9e^{-3t} + e^{-t} + e^{t} + 9e^{3t})\bigg|_{t=0} = \frac{1}{5}(9 + 1 + 1 + 9) = 4$$

6. حلول تمارين الفصل السادس

تمرين (1.6):

.1

التوزيع الاحتمالي للمتغير X

التقييم	سيئ	مقبول	جيد	جيد جدا	ممتاز
x	1	2	3	4	5
$P(x; 5)$	1/5	1/5	1/5	1/5	1/5

.2 $Var(X) = \frac{(5-1)^2}{12} = 1.33$ ، $E(X) = \frac{1+5}{2} = 3$.

.3

التوزيع الاحتمالي التراكمي للمتغير X

x	1	2	3	4	5
$F(x; 5)$	1/5	2/5	3/5	4/5	5/5

تمرين (2.6):

.1 $P(x; 20,0.45) = C_x^{20}\,(0.45)^x\,(0.55)^{20-x}$, $x = 0, 1, 2, \dots, 20$.

.2 0.16 ، .3 0.99 ، .4 المعدل = 9 نساء، والانحراف المعياري = 2.22 امرأة.

تمرين (3.6):

.1 لدينا $p = 25/500 = 0.05$ ، وبالتالي

$P(x; 7,0.05) = C_x^{7}\,(0.05)^x\,(0.95)^{7-x}$, $x = 0, 1, 2, \dots, 7$.

.2 0.004 ، .3 0.999 ، .4 0.70 ، .5 المعدل=0.35 وحدة، والتباين=0.33 وحدة.

تمرين (4.6):

$$P(1, 2, 0, 2, 0\,; 0.18\,, 0.23\,, 0.16, 0.27, 0.16\,, 5) = \frac{5!}{1!\,2!\,0!\,2!\,0!}\,.\,(0.18)^1.\,(0.23)^2.\,(0.16)^0.\,(0.27)^2.\,(0.16)^0 = 0.02$$

تمرين (5.6):

1. لدينا $P(x;p) = p\,(1-p)^{x-1} = (0.15)(0.85)^{x-1} \quad , x = 1,2,\ldots$

وبالتالي $P(X=2) = (0.15)(0.85)^{2-1} = 0.13$.

2.

$$P(X \geq 3) = 1 - P(X < 4) = 1 - \sum_{x=1}^{3} P(x; 0.15) = 1 - (0.15 + 0.13 + 0.11) = 0.61$$

3.

$$E(X) = \frac{1}{p} = \frac{1}{0.15} = \text{فترة } 6.67 \quad , \qquad Var(X) = \frac{1-p}{p^2} = \frac{1-(0.15)}{(0.15)^2} = \text{فترة } 37.78$$

تمرين (6.6):

1. لدينا $p = 0.60$ و $r = 2$ ، فيصبح شكل توزيع ذي الحدين السالب

$$P(x; 2, 0.60) = C_1^{x-1}\,(0.60)^2\,(0.40)^{x-2} \quad , x = 2, 3, 4, \ldots$$

وبالتالي $P_X(3) = C_1^{3-1}\,(0.60)^2\,(0.40)^{3-2} = 0.29$.

2. $P(X < 4) = \sum_{x=2}^{3} C_1^{x-1}\,(0.60)^2\,(0.40)^{x-2} = 0.65$.

3. $E(X) = \frac{2}{0.60} = \text{سنة } 3.33, \qquad Var(X) = \frac{2(1-0.60)}{(0.60)^2} = \text{سنة } 2.22$.

تمرين (7.6):

1. في التوزيع فوق الهندسي لدينا $N = 25$ ، $n = 3$ ، و $r = 10$ وبالتالي:

$$P(x; N, n, r) = P(x; 25, 3, 10) = \frac{C_x^{10}\,C_{3-x}^{25-10}}{C_3^{25}} \ , \quad x = 0, 1, \ldots, n$$

وهكذا يكون

$$P(3) = \frac{C_3^{10}\,C_{3-3}^{25-10}}{C_3^{25}} = 0.05$$

2.

$$P(2) = \frac{C_2^{10}\,C_{3-2}^{25-10}}{C_3^{25}} = 0.29$$

.3

$$P(X \geq 2) = \frac{C_2^{10}\ C_{3-2}^{25-10}}{C_3^{25}} + \frac{C_3^{10}\ C_{3-3}^{25-10}}{C_3^{25}} = 0.29 + 0.05 = 0.34$$

.4

$E(X) = 3 \times \frac{10}{25} =$ 1.20 مريض, $\quad Var(X) = 3 \times \frac{10}{25} \times \frac{15}{25} \times \left(\frac{25-3}{25-1}\right) =$ 0.61 مريض.

تمرين (8.6):

1. المتغير العشوائي سيتبع توزيع بواسون بالصورة

$$P(x;\ 6.5) = \frac{e^{-6.5}\ (6.5)^x}{x!} \quad , \quad x = 0, 1, 2, \dots$$

.2

$$P(X = 2) = \frac{e^{-6.5}\ (6.5)^2}{2!} = 0.03$$

.3

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.04 = 0.96$$

4. المعدل الجديد سيكون λ = 6.5 × 2 = 13 جيجا بايت، وبالتالي:

$$P(X = 2) = \frac{e^{-13}\ (13)^2}{2!} = 0.0002$$

5. $E(X) = \frac{6.5}{2} =$ 3.25 جيجا بايت , $Var(X) = \frac{6.5}{2} =$ 3.25 جيجا بايت .

تمرين (9.6):

1. الاحتمال المطلوب يمثل مساحة المستطيل (الطول × العرض) = 0.1 × 5 = 0.5 ، وهذا يعني أن 50% من الأطباق الرئيسية سيستغرق تحضيرها من 92 إلى 97 دقيقة.

2. $E(X) = \frac{90+100}{2} = 95 \quad , \quad Var(X) = \frac{(100-90)^2}{12} = 8.33$.

تمرين (10.6): نعم، ويكون

P(-1.65 < Z < 1.65) = P(-1.65 < Z < 0) + P(0 < Z < 1.65) = 0.4505 + 0.4505 = 0.9010.

تمرين (11.6):

1. المطلوب هو $P\left(X <\right.$ 24 شهر $\left.\right) = P\left(X <\right.$ 2 سنة $\left.\right)$ ، وبالتالي

$P(X < 24) = P\left(\frac{X-30}{6} < \frac{24-30}{6}\right) = P(Z < -1) = 0.1587$.

.2

$P(24 < X < 48) = P\left(\frac{24-30}{6} < Z < \frac{48-30}{6}\right) = P(-1 < Z < 3) = P(Z < 3) -$
$P(Z < -1) = 0.9987 - 0.1587 = 0.8399$.

تمرين (12.6): حيث أن المساحة تحت منحنى التوزيع الطبيعي المعياري خارج الفترة $-k < Z < k$ هي $1 - 0.95 = 0.05$ ، فيكون $P(Z < -k) = 0.05/2 = 0.025$. وهكذا فإن قيمة k التي تحقق ذلك هي 1.69.

تمرين (13.6): المطلوب هو $P(X \geq 420)$ ، ويكون

$$\mu = np = 600 \times 0.75 = 450, \;\; \sigma = \sqrt{np(1-p)} = \sqrt{600 \times 0.75 \times 0.25} = 10.61$$

وبالتالي

$$P(X \geq 420) = 1 - P(X \leq 420) = 1 - P(Z \leq -2.83) = 1 - 0.002 = 0.998.$$

تمرين (14.6):

1. $P(X \geq 0.60) = \int_{0.60}^{1} f(x;5,3)\, dx = 0.58$.

2. $E(X) = \frac{5}{5+3} = 0.625 \quad , Var(X) = \frac{5\times3}{(5+3)^2\,(5+3+1)} = 0.028$.

تمرين (15.6):

1. $P\left(X > \frac{1}{2}\right) = \int_{1/2}^{\infty} 15\, e^{-15}\, dx = 0.0006$.

2. $E(X) = \beta = 15 \quad , \quad Var(X) = \beta^2 = 225$.

7. حلول تمارين الفصل السابع

تمرين (1.7):

العينات وأوساطها الحسابية

ترتيب العينة	العينة			الوسط الحسابي للعينة	ترتيب العينة	العينة			الوسط الحسابي للعينة	ترتيب العينة	العينة			الوسط الحسابي للعينة
1	3	3	3	3.00	10	5	3	3	3.67	19	6	3	3	4.00
2	3	3	5	3.67	11	5	3	5	4.33	20	6	3	5	4.67
3	3	3	6	4.00	12	5	3	6	4.67	21	6	3	6	5.00
4	3	5	3	3.67	13	5	5	3	4.33	22	6	5	3	4.67
5	3	5	5	4.33	14	5	5	5	5.00	23	6	5	5	5.33
6	3	5	6	4.67	15	5	5	6	5.33	24	6	5	6	5.67
7	3	6	3	4.00	16	5	6	3	4.67	25	6	6	3	5.00
8	3	6	5	4.67	17	5	6	5	5.33	26	6	6	5	5.67
9	3	6	6	5.00	18	5	6	6	5.67	27	6	6	6	6.00

توزيع المعاينة للوسط

الوسط الحسابي للعينة $\bar{x}$	التكرار f	$f(\bar{x})$
3	1	0.04
3.67	3	0.11
4	3	0.11
4.33	3	0.11
4.67	6	0.22
5	4	0.15
5.33	3	0.11
5.67	3	0.11
6	1	0.04
المجموع	27	1

الوسط والتباين لتوزيع المعاينة

$$\mu_{\bar{X}} = 4.67 = \mu\,, \quad \sigma_{\bar{X}}^2 = 0.52 = \frac{1.56}{3} = \frac{\sigma^2}{n}$$

تمرين (2.7):

أولا: عندما $n = 2$:

العينات وأوساطها الحسابية

ترتيب العينة	العينة		الوسط الحسابي للعينة
1	3	5	4
2	3	6	4.5
3	5	6	5.5

توزيع المعاينة للوسط

الوسط الحسابي للعينة $\bar{x}$	التكرار f	$f(\bar{x})$
4	1	1/3
4.5	1	1/3
5.5	1	1/3
المجموع	3	1

الوسط والتباين لتوزيع المعاينة

$$\mu_{\bar{X}} = 4.67 = \mu\,, \quad \sigma_{\bar{X}}^2 = 0.39 = \frac{1.56}{3}\frac{(3-2)}{(3-1)}$$

ثانيا: عندما $n = 3$:

العينات وأوساطها الحسابية

ترتيب العينة	العينة			الوسط الحسابي للعينة
1	3	5	6	4.67

توزيع المعاينة للوسط

الوسط الحسابي للعينة $\bar{x}$	التكرار f	$f(\bar{x})$
4.67	1	1/1
المجموع	1	1

الوسط والتباين لتوزيع المعاينة

$$\mu_{\bar{X}} = 4.67 = \mu\,, \quad \sigma_{\bar{X}}^2 = 0 = \frac{1.56}{3}\frac{(3-3)}{(3-1)}$$

تمرين (3.7):

.1

$\mu_{\bar{X}} = \mu =$ 5 درجات ، $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1.5}{\sqrt{20}} =$ 0.335 درجة

.2

$$P(\bar{X} < 4) = P\left(\frac{\bar{X}-5}{0.335} < \frac{4-5}{0.335}\right) = P(Z < -2.98) = 0.001$$

.3

$$P(\bar{X} > 5.5) = P\left(\frac{\bar{X}-5}{0.335} > \frac{5.5-5}{0.335}\right) = P(Z > 1.49) = 1 - P(Z \le 1.49)$$
$$= 1 - 0.932 = 0.068$$

.4

$$P(-2.98 < Z < 1.49) = P(Z < 1.49) - P(Z < -2.98) = 0.932 - 0.001 = 0.931$$

تمرين (4.7):

.1 $\mu_1 = 5.67\,, \sigma_1^2 = 1.56$ و $\mu_2 = 4.33\,, \sigma_2^2 = 2.89$

.2

من المجتمع 1			
ترتيب العينة	العينة		الوسط الحسابي للعينة
1	4	4	4
2	4	6	5
3	4	7	5.5
4	6	4	5
5	6	6	6
6	6	7	6.5
7	7	4	5.5
8	7	6	6.5
9	7	7	7

من المجتمع 2			
ترتيب العينة	العينة		الوسط الحسابي للعينة
1	2	2	2
2	2	5	3.5
3	2	6	4
4	5	2	3.5
5	5	5	5
6	5	6	5.5
7	6	2	4
8	6	5	5.5
9	6	6	6

.3

		$\bar{X}_1$								
		7	6.5	5.5	6.5	6	5	5.5	5	4
	2	5	4.5	3.5	4.5	4	3	3.5	3	2
	3.5	3.5	3	2	3	2.5	1.5	2	1.5	0.5
	4	3	2.5	1.5	2.5	2	1	1.5	1	0
	3.5	3.5	3	2	3	2.5	1.5	2	1.5	0.5
$\bar{X}_2$	5	2	1.5	0.5	1.5	1	0	0.5	0	-1
	5.5	1.5	1	0	1	0.5	-0.5	0	-0.5	-1.5
	4	3	2.5	1.5	2.5	2	1	1.5	1	0
	5.5	1.5	1	0	1	0.5	-0.5	0	-0.5	-1.5
	6	1	0.5	-0.5	0.5	0	-1	-0.5	-1	-2

$f(\bar{X}_1 - \bar{X}_2)$	التكرار f	$\bar{X}_1 - \bar{X}_2$
0.01	1	-2
0.02	2	-1.5
0.04	3	-1
0.07	6	-0.5
0.11	9	0
0.10	8	0.5
0.12	10	1
0.15	12	1.5
0.10	8	2
0.07	6	2.5
0.10	8	3
0.05	4	3.5
0.01	1	4
0.02	2	4.5
0.01	1	5
1	81	المجموع

.4

$$\mu_{\bar{X}_1-\bar{X}_2} = 1.33 \cong 5.67 - 4.33 = \mu_1 - \mu_2$$

$$\sigma^2_{\bar{X}_1-\bar{X}_2} = 2.22 \cong \frac{1.56}{2} + \frac{2.89}{2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

تمرين (5.7):

$$\mu_{\bar{X}_{LS}-\bar{X}_{GTX}} = 70 - 50 = 20 \text{ لتر}$$

$$\sigma_{\bar{X}_{LS}-\bar{X}_{GTX}} = \sqrt{\frac{8}{100} + \frac{6}{100}} = 0.37 \text{ لتر}$$

$$P(\bar{X}_{LS} - \bar{X}_{GTX} < 19) = P\left(Z < \frac{19-20}{0.37}\right) = P(Z < -2.67) = 0.004$$

تمرين (6.7):

.1

$$\mu_p = p = 0.03 \quad , \quad \sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.03 \times 0.97}{300}} = 0.0098$$

$$P(P < 0.05) = P\left(\frac{P-\mu_p}{\sigma_p} < \frac{0.05-0.03}{0.0098}\right) = P(Z < 2.04) = 0.9793$$

.2

$$P(P > 0.03) = P\left(\frac{P-\mu_p}{\sigma_p} > \frac{0.03-0.03}{0.0098}\right) = P(Z > 0) = 1 - P(Z \le 0) = 1 - 0.5 = 0.5$$

تمرين (7.7):

.1

$$\mu_{p_1-p_2} = p_1 - p_2 = 0.05 - 0.03 = 0.02$$

$$\sigma^2_{p_1-p_2} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} = \frac{0.05(0.95)}{350} + \frac{0.03(0.97)}{350} = 0.00022$$

.2

$$P(P_1 - P_2 \le 0.04) = P\left(Z \le \frac{0.04-0.02}{0.015}\right) = P(Z \le 1.33) = 0.9082$$

تمرين (8.7):

.1 $t_{0.005}(14) = 2.977$ .2 $t_{0.01}(6) = 3.143$ .3 $t_{0.95}(24) = 1.711$

تمرين (9.7):

.1 $\chi^2_{0.05}(14) = 23.685$ .2 $\chi^2_{0.05}(14) = 48.278$ .3 $\chi^2_{0.95}(9) = 3.325$

تمرين (10.7):

.1 $f_{0.05}(5,9) = 3.48$ .2 $f_{0.01}(8,6) = 8.10$ .3 $f_{0.005}(3,7) = 10.88$.

المراجع

1. Bluman, A., (2005), *Probability Demystified*, The McGraw-Hill Companies, Inc. U.S.A.
2. Brink, D., (2008), *Statistics*, David Brink and Ventus Publishing ApS. U.S.A.
3. Douglas, M. and George, R., (2003), *Applied Statistics and Probability for Engineers*, John Wiley and Sons, Inc. U.S.A.
4. Fernandes, M., (2009), *Statistics for Business and Economics,* Marcelo Fernandes and Ventus Publishing. U.S.A.
5. Frank, H., and Althoen, S., (1994), *Statistics: Concepts and Applications,* Cambridge University Press. U.K.
6. Han, J. and Kamber, M., (2000), *Data Mining: Concepts and Techniques,* Morgan Kaufmann Publishers. U.S.A.
7. Larry, G. and Woollcott, S., (1993), *The Cartoon Guide to Statistics,* HarperCollins Publishers, Inc. U.S.A.
8. LeBlanc, D., (2004), *Statistics: Concepts and Applications for Science,* Jones and Bartlett Publishers. Canada.
9. Moore, D., (2003), *The Basic Practice of Statistics,* W. H. Freeman Publishers. U.S.A.
10. Rumsey, D., (2003), *Statistics for Dummies,* John Wiley and Sons, Inc. U.S.A.
11. Soong, T., (2004), *Fundamentals of Probability and Statistics for Engineers,* John Wiley and Sons, Inc. U.S.A.
12. Spiegel, M., Schiller, J. and Srinivasan, R., (2001), *Schaum's Easy Outline of Probability and Statistics,* The McGraw-Hill Companies, Inc. U.S.A.
13. Spiegel, M., and Stephens, L., (1999), *Theory and Problems of Statistics,* The McGraw-Hill Companies, Inc. U.S.A.
14. Stephens, L., (2006), *Schaum's Outline Beginning Statistics*, The McGraw-Hill Companies, Inc. U.S.A.
15. Wackerly, D., Mendenhall, W., and Scheaffer, R., (2002), *Mathematical Statistics with Applications,* Duxbury Thomson Learning, Inc. U.S.A.
16. Walpole, R., (1982), *Introduction to Statistics,* Macmillan Publishing Co., Inc. U.S.A.
17. Woolf, P., et al., (2004), *Statistics and Probability Primer for Computational Biologists,* Massachusetts Institute of Technology. U.S.A.